



Heriot-Watt University
Research Gateway

Optimal scaling of the MALA algorithm with irreversible proposals for Gaussian targets

Citation for published version:

Ottobre, M, Pillai, NS & Spiliopoulos, K 2019, 'Optimal scaling of the MALA algorithm with irreversible proposals for Gaussian targets', *Stochastics and Partial Differential Equations: Analysis and Computations*, pp. 1-51. <https://doi.org/10.1007/s40072-019-00147-5>

Digital Object Identifier (DOI):

[10.1007/s40072-019-00147-5](https://doi.org/10.1007/s40072-019-00147-5)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Stochastics and Partial Differential Equations: Analysis and Computations

Publisher Rights Statement:

This is a post-peer-review, pre-copyedit version of an article published in *Stochastics and Partial Differential Equations: Analysis and Computations*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s40072-019-00147-5>

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPTIMAL SCALING OF THE MALA ALGORITHM WITH IRREVERSIBLE PROPOSALS FOR GAUSSIAN TARGETS

MICHELA OTTOBRE, NATESH S. PILLAI, AND KONSTANTINOS SPILIOPOULOS

ABSTRACT. It is well known in many settings that reversible Langevin diffusions in confining potentials converge to equilibrium exponentially fast. Adding irreversible perturbations to the drift of a Langevin diffusion that maintain the same invariant measure accelerates its convergence to stationarity. Many existing works thus advocate the use of such non-reversible dynamics for sampling. When implementing Markov Chain Monte Carlo algorithms (MCMC) using time discretisations of such Stochastic Differential Equations (SDEs), one can append the discretization with the usual Metropolis-Hastings accept-reject step and this is often done in practice because the accept-reject step eliminates bias. On the other hand, such a step makes the resulting chain reversible. It is not known whether adding the accept-reject step preserves the faster mixing properties of the non-reversible dynamics. In this paper, we address this gap between theory and practice by analyzing the optimal scaling of MCMC algorithms constructed from proposal moves that are time-step Euler discretisations of an irreversible SDE, for high dimensional Gaussian target measures. We call the resulting algorithm the *ipMALA*, in comparison to the classical MALA algorithm (here *ip* is for irreversible proposal). In order to quantify how the cost of the algorithm scales with the dimension N , we prove invariance principles for the appropriately rescaled chain. In contrast to the usual MALA algorithm, we show that there could be two regimes asymptotically: (i) a diffusive regime, as in the MALA algorithm and (ii) a “fluid” regime where the limit is an ordinary differential equation. We provide concrete examples where the limit is a diffusion, as in the standard MALA, but with provably higher limiting acceptance probabilities. Numerical results are also given corroborating the theory.

Keywords: MALA Algorithm, Langevin Diffusions, Optimal Scaling, Non-reversible.

1. INTRODUCTION

In this paper, we analyze the scaling properties of high dimensional Markov Chain Monte Carlo (MCMC) algorithms constructed using non-reversible Langevin diffusions. Consider a target measure

$$\pi(dx) = \frac{1}{Z} e^{-U(x)} dx, \quad Z = \int_{\mathbb{R}^d} e^{-U(x)} dx. \quad (1.1)$$

It is known that, under mild assumptions on the potential $U(x)$, the Langevin stochastic differential equation (SDE)

$$dX_t = -\nabla U(X_t) dt + \sqrt{2} dW_t \quad (1.2)$$

has π as its unique invariant measure and is π -ergodic. For appropriate test functions $f : \mathbb{R}^d \mapsto \mathbb{R}$, by the ergodic theorem, we have

$$\frac{1}{t} \int_0^t f(X_s) ds \longrightarrow \int f(x) \pi(dx). \quad (1.3)$$

Date: July 2, 2019.

Thus the Langevin diffusion (1.1) is a fundamental tool for sampling from the target measure π or compute expectation of various functionals $\int f d\pi$, in view of (1.3). The Langevin SDE is time reversible – its generator is a self-adjoint operator in $L_2(\pi)$.

The drift of (1.2) can be modified without altering the invariant measure. Indeed, if Γ is a vector field such that $\operatorname{div}(\Gamma e^{-U}) = 0$, then diffusions of the form

$$dX_t = [-\nabla U(X_t) + \Gamma(X_t)] dt + \sqrt{2} dW_t, \quad (1.4)$$

also have π as their invariant measure. The divergence free condition can be written as

$$\operatorname{div}\Gamma = \Gamma \nabla U. \quad (1.5)$$

Observe that, for $\Gamma \neq 0$, the diffusions in (1.4) are non-reversible.

1.1. The problem. It is known that adding a non-reversible component in a Langevin SDE could accelerate its convergence to stationarity. Indications of this phenomenon are the main results of [10, 24, 25]. The main result in [10] states that, among the family of non-reversible diffusions (1.4), the one with the smallest spectral gap corresponds to $\Gamma = 0$. Similar results were established from an asymptotic variance and large deviations point of view in [24, 25]. Broadly speaking, it is a well documented principle that non-reversible dynamics have better ergodic properties than their reversible counterparts. This observation has sparked a significant amount of research work in recent years and several papers have advocated the use of non-reversible diffusions for sampling.

Numerical discretisations of (1.2) or (1.4) do not necessarily inherit the ergodic properties of the continuous-time dynamics [28]. In particular, discretised processes may not converge at all, or will have an invariant measure different from π . To circumvent these issues, in MCMC algorithms, practitioners often perform an additional Metropolis-Hastings accept-reject step for proposals constructed from time discretisations of Langevin diffusions. For instance, the standard MALA algorithm is obtained when the discretised SDE is (1.2) (see *e.g.*, [26] and [29]). To the best of our knowledge, not much is known about whether making the chain reversible by adding the accept reject step preserves the faster mixing enjoyed by the non-reversible dynamics. In this article, we seek to address this important gap between theory and practice.

1.2. Previous Work. Many different approaches were pursued in recent works in order to exploit and analyse the beneficial effects of irreversibility in the algorithmic practice of MCMC methodology. In particular: i) Irreversible algorithms have been proposed and analyzed in [2, 3, 5, 17, 13, 23] and references therein (on the matter see also [19, 20]); ii) algorithms that are obtained by discretising irreversible Markov processes in a way that the resulting Markov chain is still irreversible are studied in [9, 21]; iii) numerical algorithms that take advantage of the splitting of reversible-irreversible part of the equation are analyzed in [16, 7]. In addition, comparisons of MALA and of Langevin samplers without the accept-reject step have been performed in [6, 7]. In many cases of interest it has been observed that irreversible Langevin samplers have smaller mean square error when compared to the MALA algorithm, see [7]. The latter fact is related to the consideration that the variance reduction achieved by the irreversible Langevin sampler can be more significant than the error due to the bias of the irreversible Langevin sampler.

1.3. Our Contribution. In this paper we take a different standpoint and analyse the exact performance of a Metropolis-Hastings algorithm, on certain Gaussian target densities, where the proposal is based on discretising (1.4) as opposed to (1.2). We call this algorithm the *irreversible proposal* MALA or ipMALA for short. To quantify the efficiency of the ipMALA algorithm in high dimensions and compare it to the usual MALA algorithm, we study its optimal scaling properties and its limiting optimal acceptance probability [26, 27, 22]. Optimal scaling aims to find the “optimal size” of the local proposal variance as a function of the dimension. The optimality criteria varies for different algorithms, but a natural choice for algorithms that have a diffusion limit is the expected square jumping distance [27].

The basic mechanism for Metropolis-Hastings algorithms consists of employing a proposal transition density $q(x, y)$ in order to produce a reversible chain $\{x_k\}_{k=0}^\infty$ which has the target measure π as invariant distribution. At step k of the chain, a proposal move y_{k+1} is generated by using $q(x, y)$, *i.e.*, $y_{k+1} \sim q(x_k, \cdot)$. Then such a move is accepted with probability $\alpha(x_k, y_{k+1})$, where

$$\alpha(x_k, y_{k+1}) = \min \left\{ 1, \frac{\pi(y_{k+1})q(y_{k+1}, x_k)}{\pi(x_k)q(x_k, y_{k+1})} \right\}. \quad (1.6)$$

The MALA algorithm is a Metropolis-Hastings algorithm with proposal move generated by a time-step discretisation of the Langevin equation (1.2). The ipMALA algorithm we wish to analyze obtains proposals by discretising (1.4). As explained before, any non-trivial Γ satisfying the divergence free condition (1.5) will preserve the invariant measure. A convenient choice is to pick Γ such that

$$\operatorname{div} \Gamma = 0 \quad \text{and} \quad \Gamma \nabla U = 0. \quad (1.7)$$

A standard choice of $\Gamma(x)$ is

$$\Gamma(x) = S \nabla U(x), \quad (1.8)$$

where S is any antisymmetric matrix. A more elaborate discussion on other possible choices of $\Gamma(x)$ can be found in [24]. The meaning of the conditions (1.7) is straightforward: the flow generated by Γ must preserve Lebesgue measure since it is divergence-free; moreover, the micro-canonical measure on the surfaces $\{U = z\}$ is preserved as well.

We make one further important assumption. For the rest of the paper, we focus exclusively on Gaussian target measures. We believe most of our analysis should carry over to the important case in which the target measure has a Radon-Nikodym derivative with respect to a Gaussian measure using the methods in [18, 22].

The main result of the paper is as follows. We consider Gaussian target measures $\pi^N \sim \mathcal{N}(0, \mathcal{C}^N)$ on \mathbb{R}^N , where $\mathcal{C}^N := \operatorname{diag}\{\lambda_1^2, \dots, \lambda_N^2\}$ (see Section 2 for more details). Such a measure is clearly of the form (1.1) for a quadratic potential U . Therefore, in this case, the general form of a Euler-discretisation of (1.4) is given by

$$y_{k+1}^N = x_k^N - \frac{\sigma_N^2}{2} x_k^N + \sigma_N^\alpha \mathcal{C}^N S^N x_k^N + \sigma_N (\mathcal{C}^N)^{1/2} z_{k+1}^N,$$

where $z_{k+1}^N \sim \mathcal{N}(0, \mathbf{I}_N)$, $\sigma_N = \ell/N^\gamma$ and $\alpha > 0$. The notation S^N is used to stress that here S^N is an $N \times N$ antisymmetric matrix. The quantity y_{k+1}^N is the proposal, and then a Markov chain is formed using the usual accept-reject mechanism (more details about the algorithm can be found in Section 3). Let x_k^N be the resulting ipMALA Markov chain. Interestingly,

depending on whether α is bigger, equal or smaller than two, we will have different limiting behaviours. Broadly speaking, we show the following:

- if $\gamma < 1/6$ then the acceptance probability degenerates to zero exponentially quickly, and thus this case is not of practical interest.
- if $\alpha \geq 2$ then the optimal value of γ is $\gamma = 1/6$. We prove that the continuous interpolant of two subsequent steps of the MALA algorithm, see (4.6) for proper definition, has a diffusion limit. The cost of the algorithm is still $N^{1/3}$ like in the standard MALA case. However, if S^N and α are chosen appropriately (see specific examples in Sections 6 and 7), the limiting acceptance probability is higher. That is, the ipMALA will accept moves more frequently than MALA.
- if $1 \leq \alpha < 2$ then take $\gamma \geq 1/6$. In this regime we show that the cost of the algorithm is of the order $N^{\alpha\gamma}$, and one can choose $\alpha\gamma < 1/3$. In addition, we show that the continuous interpolant, see (4.6), converges weakly to the solution of a deterministic ODE. It is also interesting that the ODE that we get in the limit in this regime, can be related to Hamilton's ODE in HMC (see Theorem 4.2) and we plan to investigate this in the future.

We would like to stress that in this paper we only study different scaling regimes when the algorithm is started *in stationarity*. In the case of the MALA algorithm it is a known fact that the optimal scaling out of stationarity differs from the optimal scaling in stationarity (the former being $O(N^{1/2})$ and the latter being $O(N^{1/3})$, see [4, 11, 12, 15]). It would be relevant to address the same issue for the ipMALA algorithm (especially in view of the fact that non-reversibility is known to speed up convergence, although we do realise that the measure of efficiency we use here may well be unrelated to convergence rate) but we do not do it in this paper, as this would involve substantial further analysis.

The goal of this paper is to explore in a rigorous mathematical way what happens when a Metropolis-Hastings accept-reject step is applied to a proposal coming from discretisation of an irreversible Langevin diffusion. We find that there are different possible regimes and we characterize explicitly what is the limit of the continuous interpolant of the chain as the number of steps goes to ∞ , see Theorem 4.2. We find that the irreversible perturbation matrix has non-trivial effects on the limiting dynamics (see Theorem 4.2) and on the limiting average optimal acceptance probability, see (5.29) and Section 6. In terms of applications, the conclusion is that even though the introduction of the accept-reject step potentially offsets some of the advantages of irreversible perturbations [10, 24, 25], some advantages may appear in certain cases. In particular, by appropriately choosing S and α (e.g. $S = S_1$ and $\alpha > 4$ as in Section 6), we can obtain a diffusion limit but with limiting acceptance probabilities higher than that of MALA. In addition, even in the fluid regime, where the limit of the algorithm is an ODE, the acceptance probability can be the same as that of MALA, but the algorithm takes, in stationarity, $N^{\alpha\gamma}$ steps to explore the state space with $\alpha\gamma < 1/3$ as opposed to $N^{1/3}$ steps for standard MALA.

Lastly, it is possible that for multi-modal targets the use of irreversible proposals may be beneficial as this is the case in the absence of the accept-reject step, see [25]; an analysis of this fact is though beyond the scope of this paper. The present paper certainly invites further research on this topic in different directions, including multi-modal targets, out of stationarity analysis and computational considerations of implementation issues.

The paper is organized as follows. In Section 2 we introduce the notation used throughout the paper. In Section 3 we describe and motivate the algorithm that we will examine. Section 4 contains the rigorous statement of our main result and its main implications. In Section 5 we give a detailed heuristic argument to explain how the main result is obtained and why it should hold. In Section 6 we present some examples of potential choices for the antisymmetric matrix and the computation of the corresponding limiting acceptance probabilities. Section 7 contains extensive simulation studies that demonstrate the theoretical results. Rigorous proofs are relegated to the Appendix.

2. PRELIMINARIES AND NOTATION

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$ denote an infinite dimensional separable Hilbert space with the canonical norm derived from the inner-product. Let \mathcal{C} be a positive, trace class operator on \mathcal{H} and $\{\phi_j, \lambda_j^2\}_{j \geq 1}$ be the eigenfunctions and eigenvalues of \mathcal{C} respectively, so that

$$\mathcal{C}\phi_j = \lambda_j^2 \phi_j, \quad \sum_j \lambda_j^2 < \infty. \quad (2.1)$$

The eigenvalues λ_j^2 are non-decreasing. We assume a normalization under which $\{\phi_j\}_{j \geq 1}$ forms a complete orthonormal basis in \mathcal{H} . Let π be a Gaussian probability measure on \mathcal{H} with covariance operator \mathcal{C} , that is,

$$\pi \sim \mathcal{N}(0, \mathcal{C}).$$

If X^N is the finite dimensional space

$$\mathcal{H} \supset X^N := \text{span}\{\phi_j\}_{j=1}^N$$

spanned by the first N eigenvectors of the covariance operator (notice that the space X^N is isomorphic to \mathbb{R}^N), then for any fixed $N \in \mathbb{N}$ and $x \in \mathcal{H}$ we denote by $\mathcal{P}^N(x)$ the projection of x on X^N . Moreover, the finite dimensional projection π^N of the measure π on X^N is given by the Gaussian measure

$$\pi^N \sim \mathcal{N}(0, \mathcal{C}^N)$$

where $\mathcal{C}^N := \text{diag}\{\lambda_1^2, \dots, \lambda_N^2\}$ (or, more precisely, $\mathcal{C}^N := \mathcal{P}^N \circ \mathcal{C} \circ \mathcal{P}^N$). Thus we have,

$$\pi^N := \pi(x^N) = \pi(x^{1,N}, \dots, x^{N,N}) = \frac{1}{(\sqrt{2\pi})^N} \prod_{i=1}^N \frac{1}{\lambda_i} e^{-\frac{|x^{i,N}|^2}{2\lambda_i^2}}. \quad (2.2)$$

Let $\tilde{S} : \mathcal{H} \rightarrow \mathcal{H}$ be a bounded linear operator. Thus there exists a constant $\kappa > 0$ such that

$$\|\tilde{S}x\| \leq \kappa \|x\|, \quad x \in \mathcal{H}. \quad (2.3)$$

For any $N \in \mathbb{N}$ we can consider the projected operator \tilde{S}^N , defined as follows:

$$\tilde{S}^N x := (\mathcal{P}^N \circ \tilde{S} \circ \mathcal{P}^N)x.$$

The operator \tilde{S}^N is also bounded on \mathcal{H} . Since X^N is isomorphic to \mathbb{R}^N , \tilde{S}^N can be represented by an $N \times N$ matrix. Throughout the paper we require the following: \tilde{S} is such that for any $N \in \mathbb{N}$, the matrix \tilde{S}^N can be expressed as the product of a symmetric matrix, namely \mathcal{C}^N , and an antisymmetric matrix, S^N :

$$\tilde{S}^N = \mathcal{C}^N S^N, \quad \text{for every } N \in \mathbb{N}. \quad (2.4)$$

Throughout the paper we will use the following notation:

- x and y are elements of the Hilbert space \mathcal{H} ;
- the letter N is reserved to denote the dimensionality of the space X^N where the target measure π^N is supported;
- x^N is an element of $X^N \cong \mathbb{R}^N$ (similarly for y^N and the noise ξ^N); the j -th component of the N -dimensional vector x^N (in the basis $\{\varphi_j\}_{j=1}^N$) is denoted by $x^{j,N}$.

We analyse Markov chains evolving in \mathbb{R}^N . Because the dimensionality of the space in which the chain evolves will be a key fact, we want to keep track of both the dimension N and the step k of the chain. Therefore,

- x_k^N will denote the k -th step of the chain $\{x_k^N\}_{k \in \mathbb{N}}$ evolving in \mathbb{R}^N ;
- compatibly with the notation set above, $x_k^{i,N}$ is the i -th component of the vector $x_k^N \in \mathbb{R}^N$; i.e., $x_k^{i,N} = \langle x_k^N, \varphi_i \rangle$;
- two (double) sequences of real numbers $\{A_k^N\}$ and $\{B_k^N\}$ satisfy $A_k^N \lesssim B_k^N$ if there exists a constant $K > 0$ (independent of N and k) such that

$$A_k^N \leq K B_k^N,$$

for all N and k such that $\{A_k^N\}$ and $\{B_k^N\}$ are defined. The notation \lesssim is used for functions and random functions as well, with the constant K being independent of the argument of the function and of the randomness.

As we have already mentioned, $\|\cdot\|$ is the norm on \mathcal{H} , namely

$$\|x\|^2 := \sum_{i=1}^{\infty} \langle x, \varphi_i \rangle^2.$$

With abuse of notation, we will also write

$$\|x^N\|^2 = \sum_{i=1}^N |x^{i,N}|^2.$$

We will also use the weighted norm $\|\cdot\|_{\mathcal{C}}$, defined as follows:

$$\|x\|_{\mathcal{C}}^2 := \langle x, \mathcal{C}^{-1}x \rangle = \sum_{i=1}^{\infty} \frac{|\langle x, \varphi_i \rangle|^2}{\lambda_i^2},$$

for all $x \in \mathcal{H}$ such that the above series is convergent; analogously,

$$\|x^N\|_{\mathcal{C}^N}^2 := \sum_{i=1}^N \frac{|x^{i,N}|^2}{\lambda_i^2}.$$

If A^N is an $N \times N$ matrix, $A^N = \{A_{i,j}^N\}$, we define

$$\begin{aligned} \mathcal{V}_A^N &:= \sum_{i=1}^N \lambda_i^2 \sum_{j=1}^N |A_{i,j}^N|^2 \lambda_j^2 \\ &= \mathbb{E}_{\pi^N} \|(\mathcal{C}^N)^{1/2} A^N x^N\|^2 = \mathbb{E}_{\pi^N} \left| \langle (\mathcal{C}^N)^{1/2} z^N, A^N x^N \rangle \right|^2 \end{aligned} \quad (2.5)$$

In the above $z^N \sim \mathcal{N}(0, I_N)$, $x^N \sim \pi^N$ and \mathbb{E}_{π^N} denotes expectation with respect to all the sources of noise contained in the integrand. We will often also write \mathbb{E}_k and \mathbb{E}_x , to mean

$$\mathbb{E}_k = \mathbb{E}[\cdot | x_k], \quad \mathbb{E}_x = \mathbb{E}[\cdot | x_k = x].$$

Finally, many of the objects of interest in this paper depend on the matrix S^N and on the parameters α and γ . When we want to stress such a dependence, we will add a subscript \mathbf{p} , see for example the notation for the drift $d_{\mathbf{p}}$ in Theorem 4.2.

3. THE IPMALA ALGORITHM

As we have already mentioned, the classical MALA algorithm is a Metropolis-Hastings algorithm which employs a proposal that results from a one-step discretisation of the Langevin dynamics (1.2). This is motivated by the fact that the dynamics (1.2) is ergodic with unique invariant measure given by π (1.1) and can therefore be used to sample from π . The Langevin dynamics that samples from our target of interest, the Gaussian measure π^N in (2.2), reads as follows:

$$dX_t = -(\mathcal{C}^N)^{-1} X_t dt + \sqrt{2} dW_t, \quad X_t \in X^N,$$

where W_t is a N -dimensional standard Wiener process. If Δ is any positive definite, N -dimensional symmetric matrix, then the equation

$$dX_t = -\Delta(\mathcal{C}^N)^{-1} X_t dt + \sqrt{2\Delta} dW_t$$

is still ergodic with invariant measure π^N in (2.2). Now notice that if \mathcal{C} is a trace class operator, then \mathcal{C}^{-1} is unbounded, so in the limit as $N \rightarrow \infty$ any of the above two dynamics (and their discretisations) would lead to numerical instabilities. To avoid the appearance of unbounded operators, we can choose $\Delta = \mathcal{C}^N/2$ and therefore consider the SDE

$$dX_t = -\frac{1}{2} X_t dt + (\mathcal{C}^N)^{1/2} dW_t.$$

Discretizing the above and using such a discretisation as a Metropolis-Hastings proposal would result in a well-posed MALA algorithm to sample from (2.2). However, as explained in the Introduction, here we want to analyze the MALA algorithm with irreversible proposal. As in (1.8), we next consider the non-reversible SDE:

$$dX_t = \left(-\frac{1}{2} X_t + S^N (\mathcal{C}^N)^{-1} X_t \right) dt + (\mathcal{C}^N)^{1/2} dW_t,$$

where S^N is any $N \times N$ antisymmetric matrix. Again to avoid the appearance of unbounded operators, we modify the irreversible part of drift term and, finally, obtain the dynamics

$$dX_t = \left(-\frac{1}{2} X_t + \mathcal{C}^N S^N X_t \right) dt + (\mathcal{C}^N)^{1/2} dW_t. \quad (3.1)$$

Notice that, for any $x^N \in X^N$,

$$\nabla \cdot ((\mathcal{C}^N S^N x^N) \pi^N) = \text{Trace}(\mathcal{C}^N S^N) \pi^N - \langle \mathcal{C}^N S^N x^N, (\mathcal{C}^N)^{-1} x^N \rangle \pi^N = 0, \quad (3.2)$$

having used the antisymmetry of S^N and the symmetry of \mathcal{C}^N . Therefore, π^N is invariant for the dynamics (3.1). This justifies using the Metropolis-Hastings proposal (3.3) below, which is a one-step Euler discretisation of the SDE (3.1).

We now describe the ipMALA algorithm. If the chain is in x_k^N at step k , the ipMALA algorithm has the proposal move:

$$y_{k+1}^N = x_k^N - \frac{\sigma_N^2}{2} x_k^N + \sigma_N^\alpha \mathcal{C}^N S^N x_k^N + \sigma_N (\mathcal{C}^N)^{1/2} z_{k+1}^N \quad (3.3)$$

where

$$\sigma_N = \frac{\ell}{N^\gamma}, \quad \ell, \gamma > 0 \quad (3.4)$$

and $\alpha > 0$. The proposal (3.3) is a generalised discretisation of (3.1). We may indeed choose α so that, asymptotically, relative to the reversible drift and diffusion, the non-reversible drift dominates ($1 \leq \alpha < 2$), vanishes ($\alpha > 2$) or is balanced ($\alpha = 2$). This will result in different scaling limits. The choice of the parameters α, γ and ℓ will be further discussed in Section 4 below. S^N can be any antisymmetric $N \times N$ matrix and $z_{k+1}^N = (z_{k+1}^{1,N}, \dots, z_{k+1}^{N,N})$ is a vector of i.i.d standard Gaussians. With this proposal, the acceptance probability is

$$\beta^N(x^N, y^N) = 1 \wedge e^{Q^N(x^N, y^N)} \quad (3.5)$$

where

$$\begin{aligned} Q^N(x^N, y^N) = & -\frac{\sigma_N^2}{8} (\|y^N\|_{\mathcal{C}^N}^2 - \|x^N\|_{\mathcal{C}^N}^2) + \frac{1}{2} \sigma_N^{2\alpha-2} (\|\mathcal{C}^N S^N x^N\|_{\mathcal{C}^N}^2 - \|\mathcal{C}^N S^N y^N\|_{\mathcal{C}^N}^2) \\ & + 2\sigma_N^{\alpha-2} \langle S^N y^N, x^N \rangle. \end{aligned}$$

The above expression for Q^N is obtained with straightforward calculations, after observing that the proposal kernel $q(x^N, y^N)$ implied by (3.3) is such that

$$q(x^N, y^N) \propto \exp \left\{ -\frac{1}{2\sigma_N^2} \|y^N - x^N + \frac{\sigma_N^2}{2} x^N - \sigma_N^\alpha \mathcal{C}^N S^N x^N\|_{\mathcal{C}^N}^2 \right\}.$$

If $\tilde{\beta}^N \sim \text{Bernoulli}(\beta^N(x^N, y^N))$, then the Metropolis-Hastings chain $\{x_k^N\}_{\{k \in \mathbb{N}\}}$ resulting from using the proposal (3.3) can be written as follows

$$x_{k+1}^N = \tilde{\beta}^N y_{k+1}^N + (1 - \tilde{\beta}^N) x_k^N = x_k^N + \tilde{\beta}^N (y_{k+1}^N - x_k^N). \quad (3.6)$$

4. MAIN RESULTS AND IMPLICATIONS

To understand the behaviour of the chain (3.6), we start by decomposing it into its drift and martingale part; that is, let $\zeta > 0$ and write

$$x_{k+1}^N = x_k^N + \frac{1}{N^{\zeta\gamma}} d_{\mathbf{p}}^N(x_k^N) + \frac{1}{N^{\zeta\gamma/2}} M_{k,\mathbf{p}}^N \quad (4.1)$$

where the approximate drift $d_{\mathbf{p}}^N$ is defined as

$$d_{\mathbf{p}}^N(x_k^N) := N^{\zeta\gamma} \mathbb{E}_k(x_{k+1}^N - x_k^N) \quad (4.2)$$

and the approximate diffusion $M_{k,\mathbf{p}}^N$ is given by

$$M_{k,\mathbf{p}}^N := N^{\zeta\gamma/2} [x_{k+1}^N - x_k^N - \mathbb{E}_k(x_{k+1}^N - x_k^N)].$$

Using (3.3)-(3.4) and (3.6), we rewrite the approximate drift (4.2) as follows:

$$d_{\mathbf{p}}^N(x_k^N) = N^{\zeta\gamma} \mathbb{E}_k \left[(1 \wedge e^{Q^N}) \left(-\frac{\ell^2}{2N^{2\gamma}} x_k^N + \frac{\ell^\alpha}{N^{\alpha\gamma}} \mathcal{C}^N S^N x_k^N \right) \right] \quad (4.3)$$

$$+ N^{\zeta\gamma} \mathbb{E}_k \left[(1 \wedge e^{Q^N}) \frac{\ell}{N^\gamma} (\mathcal{C}^N)^{1/2} z_{k+1}^N \right]. \quad (4.4)$$

Looking at (4.3) it is clear that we need to examine two different cases, namely

- i) **Diffusive regime** : $2\gamma \leq \alpha\gamma$, i.e., $\alpha \geq 2$
- ii) **Fluid regime** : $2\gamma > \alpha\gamma$, i.e., $\alpha < 2$.

The names of the above regimes will be clear after the statement of our main results, Theorem 4.1 and Theorem 4.2, see also Remark 4.4. We will show that, in order for the algorithm to have a well defined non-trivial limit, in the case i) we should choose $\zeta = 2$, whereas in case ii) we should choose $\zeta = \alpha$. For this reason, it is intended from now on that

$$\zeta = \alpha \quad \text{if } \alpha < 2 \quad \text{and} \quad \zeta = 2 \quad \text{if } \alpha \geq 2. \quad (4.5)$$

With this observation in mind, we introduce the continuous interpolant of the chain $\{x_k^N\}$:

$$x^{(N)}(t) = (N^{\zeta\gamma}t - k)x_{k+1}^N + (k + 1 - N^{\zeta\gamma}t)x_k^N, \quad t_k \leq t < t_{k+1}, \quad (4.6)$$

where $t_k = k/N^{\zeta\gamma}$.

We now come to state the main results of this paper, Theorem 4.1 and Theorem 4.2. Theorem 4.2 is the most general result. The assumptions under which Theorem 4.2 holds are a bit involved; we detail and motivate them in Section 5.1. In this section we first state Theorem 4.1, which is just Theorem 4.2, adapted to the case in which the sequence of matrices S^N is chosen from a specific class. Namely, we restrict our attention to the case in which each of the matrices S^N is in Jordan block form: we assume that $S_{i,j}^N = J_i$ and $S_{j,i}^N = -J_i$ for $j = i + 1$ and $i = 1, 3, 5, 7, \dots, N - 1$ and $S_{i,j}^N = 0$ otherwise; here $\{J_i\}_i$ is an arbitrary sequence of real numbers. To be more clear, when we refer to matrices S^N in Jordan block form, we are referring to matrices of the form

$$\begin{pmatrix} 0 & J_1 & 0 & 0 & 0 & 0 \\ -J_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & J_2 & 0 & 0 \\ 0 & 0 & -J_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & \end{pmatrix}. \quad (4.7)$$

In this case, we can carry out explicit calculations, construct examples and demonstrate our results.

Theorem 4.1. *Assume that the family of anti-symmetric matrices S^N is in Jordan block form (i.e. of the form 4.7 as discussed above) and $x_0 \sim \pi$. Assume the constant*

$$c_1 = \lim_{N \rightarrow \infty} \frac{\mathbb{E}_{\pi^N} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2}{N^{2(\alpha-1)\gamma}} \quad \text{is finite.} \quad (4.8)$$

Then, as $N \rightarrow \infty$, the continuous interpolant $x^{(N)}(t)$ of the chain $\{x_k^N\}$ (defined in (4.6)-(4.5) and (3.6), respectively) started from the state $x_0^N = \mathcal{P}^N(x_0)$, converges weakly in $C([0, T]; \mathcal{H})$ to the solution of the following equation

$$dx(t) = d_{\mathbf{p}}(x(t))dt + D_{\mathbf{p}} dW_t^{\mathcal{C}}, \quad x(t) \in \mathcal{H}, \quad (4.9)$$

where $W_t^{\mathcal{C}}$ is a \mathcal{H} -valued Brownian motion with covariance \mathcal{C} . The drift coefficient $d_{\mathbf{p}}(x) : \mathcal{H} \rightarrow \mathcal{H}$ and the diffusion constant $D_{\mathbf{p}} \geq 0$ are defined as follows:

$$d_{\mathbf{p}}(x) := \begin{cases} -\frac{\ell^2}{2} h_{\mathbf{p}}^J x & \text{if } \gamma = 1/6 \text{ and } \alpha \geq 2 \\ 0 & \text{if } \gamma \geq 1/6 \text{ and } 1 \leq \alpha < 2, \end{cases} \quad (4.10)$$

and

$$D_{\mathbf{p}} := \begin{cases} \ell \sqrt{h_{\mathbf{p}}^J} & \text{if } \gamma = 1/6 \text{ and } \alpha \geq 2 \\ 0 & \text{if } \gamma \geq 1/6 \text{ and } 1 \leq \alpha < 2. \end{cases} \quad (4.11)$$

The real constant $h_{\mathbf{p}}^J$ is defined as

$$h_{\mathbf{p}}^J := 2\Phi \left(\frac{-\ell^6/32 - a}{\sqrt{(\ell^6/16) + 2a}} \right), \quad (4.12)$$

where $a = 2\ell^{2(\alpha-1)}c_1$ and Φ is the standard Gaussian cumulative distribution function. Lastly, as measured by the number of steps, in stationarity, to explore the state space the cost of the algorithm in the diffusion case, $\alpha \geq 2$, is $N^{1/3}$ whereas in the fluid case, $1 \leq \alpha < 2$, it is $N^{\alpha\gamma}$.

We now move on to stating Theorem 4.2. The precise statement of the assumptions of the theorem is deferred to Section 5. After stating Theorem 4.2, we first compare it with Theorem 4.1 (Remark 4.3) and then make several observations regarding practical implications of the statement (Remark 4.4). We stress that the operator \tilde{S} appearing in the statement of the theorem is as in Section 2 (see (2.3)-(2.4)). In particular, the projected operator \tilde{S}^N is, for every N , the product of the matrix \mathcal{C}^N and of a generic antisymmetric matrix S^N , appearing in the proposal (3.3).

Theorem 4.2. *Let Assumption 5.1, Assumption 5.2, Condition 5.3 and Assumption 5.4 hold and let $x_0 \sim \pi$. Then, as $N \rightarrow \infty$, the continuous interpolant $x^{(N)}(t)$ of the chain $\{x_k^N\}$ (defined in (4.6)-(4.5) and (3.6), respectively) started from the state $x_0^N = \mathcal{P}^N(x_0)$, converges weakly in $C([0, T]; \mathcal{H})$ to the solution of the following equation*

$$dx(t) = d_{\mathbf{p}}(x(t))dt + D_{\mathbf{p}} dW_t^{\mathcal{C}}, \quad x(t) \in \mathcal{H}, \quad (4.13)$$

where $W_t^{\mathcal{C}}$ is a \mathcal{H} -valued Brownian motion with covariance \mathcal{C} . In addition, the drift coefficient $d_{\mathbf{p}}(x) : \mathcal{H} \rightarrow \mathcal{H}$ and the diffusion constant $D_{\mathbf{p}} \geq 0$ are defined as follows:

$$d_{\mathbf{p}}(x) := \begin{cases} -\frac{\ell^2}{2} h_{\mathbf{p}} x & \text{if } \gamma = 1/6 \text{ and } \alpha > 2 \\ -\frac{\ell^2}{2} h_{\mathbf{p}} x + \tau_{\mathbf{p}} \ell^2 \tilde{S} x & \text{if } \gamma = 1/6 \text{ and } \alpha = 2 \\ \tau_{\mathbf{p}} \ell^{\alpha} \tilde{S} x & \text{if } \gamma \geq 1/6 \text{ and } 1 \leq \alpha < 2, \end{cases} \quad (4.14)$$

and

$$D_{\mathbf{p}} := \begin{cases} \ell \sqrt{h_{\mathbf{p}}} & \text{if } \gamma = 1/6 \text{ and } \alpha \geq 2 \\ 0 & \text{if } \gamma \geq 1/6 \text{ and } 1 \leq \alpha < 2 \end{cases} \quad (4.15)$$

The real constants $h_{\mathbf{p}}$ and $\tau_{\mathbf{p}}$ are defined in (5.28) and in (5.32) respectively.

Proof. See Appendix A and B. □

Remark 4.3. If the sequence of matrices S^N is in Jordan block form, then all the assumptions of Theorem 4.2 are satisfied, provided (4.8) holds.¹ It should be clearly said that, while examples and calculations in the case of Theorem 4.1 are very explicit, we have not been able to find examples of matrices S^N such that the constant $\tau_{\mathbf{p}}$ in Theorem 4.2 is non-vanishing.

However, it is important to stress that even when $\tau_{\mathbf{p}} = 0$ it is still possible that the average acceptance probability (given, in the limit, by $h_{\mathbf{p}}$) is not tending to one and it is of order 1 instead, see the examples in Sections 6 and 7. Notice that $\tau_{\mathbf{p}} = 0$ when $b = 2a$, with b and a are parameters depending on the choice of S^N , α and γ , defined in (5.26) and (5.27). Cfr Table 2 and Table 1 to have a summary of how $h_{\mathbf{p}}$ and $\tau_{\mathbf{p}}$ depend on a and b . As the numerical studies of Sections 6 and 7 demonstrate, even in the setup of Theorems 4.1, for which we can construct practical algorithms, we end up with interesting non-trivial results.

We now comment on the practical implications of Theorems 4.1 and 4.2.

Remark 4.4. The limiting drifts (and corresponding diffusion coefficients) appearing in (4.14) correspond to the two different regimes i) and ii), which we identified before the statement of the theorem. The choice $\gamma = 1/6$ in case (i) (*i.e.*, in the diffusive regime) and $\gamma \geq 1/6$ in the fluid regime will be fully motivated in Section 5.1. The various results obtained in the different regimes are summarised in Tables 1 and 2 below.

- In the regime (i), the effective time-step implied by the interpolation (4.5)–(4.6), is $N^{-2\gamma} = N^{-1/3}$. Therefore, if $\gamma = 1/6$ and $\alpha \geq 2$, Theorem 4.2 implies that the optimal scaling for the proposal variance when the chain is in its stationary regime is of the order $N^{-1/3}$. That is, the cost of the algorithm (in terms of number of steps needed, in stationarity, to explore the state space) is $O(N^{1/3})$. This is the same scaling as obtained in [26, 22] for the MALA algorithm. Therefore, in the regime (i), the ipMALA algorithm has the same scaling properties of the usual MALA. More discussions on this can be found in Section 5.2.
- In the case $\gamma = 1/6$ and $\alpha > 2$, one can construct specific matrices S , which result in considerably higher limiting acceptance probabilities than in the classical MALA case, see Sections 6 and 7.
- When $\alpha < 2$ and $\gamma \geq 1/6$, the (rescaled) chain converges to a fluid limit; *i.e.*, the limiting behaviour of the chain is described by an ODE. Such an ODE still admits our target as invariant measure. With the same reasoning as above, the main result implies that, in this regime, the cost of the algorithm is of the order $N^{\alpha\gamma}$ and one can choose $\alpha\gamma < 1/3$. In the fluid regime, one may expect that the chain gets to get stuck in the limit as $N \rightarrow \infty$. However, we also comment here that in the simulations that we performed we did not see the chain getting stuck in this regime, which may imply that one may need to go to very high dimensions to observe such issues.

¹Indeed, in this case Assumption 5.1 and Assumption 5.2 presented in Section 5 are satisfied with c_1 as in (4.8) and $c_2 = c_3 = 0$. This makes Assumption 5.4 easy to verify. Moreover, Condition 5.3 is trivially satisfied for matrices in Jordan block form. Detailed comments on this can be found in Section 5, see comments after (5.29).

Setting of Theorem 4.1: $c_j = 0$ for $j = 2, 3$ and $b = 2a$		
$\gamma = 1/6 \Rightarrow$ acc. prob is $O(1)$		
$\alpha \geq 2$	$dx_t = -\frac{\ell^2}{2} h_{\mathbf{p}}^J x_t dt + \ell \sqrt{h_{\mathbf{p}}^J} dW_t^c$	$N^{1/3}$
$1 < \alpha < 2$	$dx_t = 0$	$N^{\alpha\gamma}$
$\gamma > 1/6 \Rightarrow$ acc. prob tends to one		
$\alpha \geq 2$	too costly $N^{\alpha\gamma} (> N^{1/3})$	
$1 \leq \alpha < 2$	$dx_t = 0$	$N^{\alpha\gamma}$

TABLE 1. The possible different regimes when the antisymmetric matrix S takes the Jordan block form. The constants c_j are those appearing in Assumption 5.2; such constants also determine the value of the parameters a and b (defined after (5.25))

Setting of Theorem 4.2: case $c_j \neq 0$ for $j = 1, 2, 3$ and $b \neq 2a$		
$\gamma = 1/6 \Rightarrow$ acc. prob is $O(1)$		
$\alpha > 2$	$dx_t = -\frac{\ell^2}{2} h_{\mathbf{p}} x_t dt + \ell \sqrt{h_{\mathbf{p}}} dW_t^c$	$N^{1/3}$
$\alpha = 2$	$dx_t = -\frac{\ell^2}{2} h_{\mathbf{p}} x_t dt + \tau_{\mathbf{p}} \ell^2 \tilde{S} x_t dt + \ell \sqrt{h_{\mathbf{p}}} dW_t^c$	$N^{1/3}$
$1 \leq \alpha < 2$	$dx_t = \tau_{\mathbf{p}} \ell^\alpha \tilde{S} x_t dt$	$N^{\alpha\gamma}$
$\gamma > 1/6 \Rightarrow$ acc. prob is $O(1)$		
$\alpha \geq 2$	too costly $N^{\alpha\gamma} (> N^{1/3})$	
$1 \leq \alpha < 2$	$dx_t = \tau_{\mathbf{p}} \ell^\alpha \tilde{S} x_t dt$	$N^{\alpha\gamma}$

TABLE 2. The above table summarizes the general scenario described by Theorem 4.2 in the case in which $c_j \neq 0$ for $j = 1, 2, 3$ and $b \neq 2a$. The constants c_j are those appearing in Assumption 5.2; such constants also determine the value of the parameters a and b (defined after (5.25)). Table 1 is a subcase of this table. Indeed, Theorem 4.2 also covers the following: if $b = 2a$ and $c_j \neq 0$ for at least one j then all of the above holds with $\tau_{\mathbf{p}} = 0$. If $b = 2a$ and all the c_j 's vanish, then all of the above holds with $\tau_{\mathbf{p}} = 0$.

5. HEURISTIC DERIVATION OF THE DIFFUSION LIMIT

In this section we give heuristic arguments to explain how one can formally obtain the diffusion limit of Theorem 4.2 for the chain $\{x_k^N\}_k$. We stress that the arguments of this section are only formal; therefore, we often use the notation “ \simeq ”, to mean “approximately equal”. We write $A \simeq B$ when $A = B +$ “terms that are negligible” as N tends to infinity; we then rigorously justify these approximations, and the resulting limit theorems, in the Appendix.

The goal of this section is to motivate (a): the proofs of Theorems 4.1 and 4.2 that are presented in the Appendix A and B, and (b): the Assumptions 5.1-5.4 needed for the general limiting result to hold.

In particular, in Subsection 5.1, we present the heuristic derivation and the intuition behind it for the asymptotic acceptance probability as the dimension increases. On the way of doing so, we present and justify the assumptions that need to be imposed in order for

the appropriate scaling limits to hold. In addition, Remark 5.5 discusses at length these assumptions and presents examples where they are expected to hold. Subsections 5.2 and 5.3 present the heuristic calculations and motivation for the drift and diffusion coefficient respectively of the limiting continuous time interpolation of the resulting Markov chain. The heuristic discussion of Subsections 5.1, 5.2 and 5.3 is made rigorous in Appendix A and B.

5.1. Study of the acceptance probability and statement of main assumptions. In order to understand the behaviour of the chain, it is crucial to gain intuition about the acceptance probability β^N . While attempting to improve such an intuition in this section, we also present the motivation behind the assumptions needed for Theorem 4.2.

Using (3.3), a more useful and detailed expression for Q^N (which was introduced just after (3.5)) is

$$Q^N(x^N, y^N) = \bar{Q}^N(x^N, y^N) + Q_\alpha^N(x^N, y^N) \quad (5.1)$$

where \bar{Q}^N contains all the terms that come from the reversible part of the proposal while Q_α^N contains all the terms that come from the irreversible part of the proposal; namely,

$$\begin{aligned} \bar{Q}^N := & \left(-\frac{\sigma_N^3}{4} + \frac{\sigma_N^5}{8} \right) \langle x^N, (\mathcal{C}^N)^{1/2} z^N \rangle_{\mathcal{C}^N} + \frac{\sigma_N^4}{8} (\|x^N\|_{\mathcal{C}^N}^2 - \|z^N\|^2) \\ & - \frac{\sigma_N^6}{32} \|x^N\|_{\mathcal{C}^N}^2 \end{aligned}$$

and

$$Q_\alpha^N := \frac{1}{2} \sigma_N^{2\alpha} \left(\|\tilde{S} x^N\|_{\mathcal{C}^N}^2 - \|\tilde{S}^N (\mathcal{C}^N)^{1/2} z^N\|_{\mathcal{C}^N}^2 \right) - \left(2\sigma_N^{\alpha-1} + \frac{1}{4} \sigma_N^{3+\alpha} \right) \langle (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle \quad (5.2)$$

$$- \left(2\sigma_N^{2\alpha-2} + \frac{1}{4} \sigma_N^{2\alpha+2} \right) \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 - \left(\sigma_N^{2\alpha-1} - \frac{1}{2} \sigma_N^{2\alpha+1} \right) \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N} \quad (5.3)$$

$$- \sigma_N^{3\alpha-1} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} - \frac{1}{2} \sigma_N^{4\alpha-2} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2 \quad (5.4)$$

$$- \sigma_N^{3\alpha-2} \left(1 - \frac{\sigma_N^2}{2} \right) \langle \tilde{S}^N x^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N}. \quad (5.5)$$

For N large, from (3.4) we have $\sigma_N^{\alpha-1} \gg \sigma_N^{3+\alpha}$; so we expect that, asymptotically, the last term in (5.2) will be negligible. The same reasoning can be applied to the terms in (5.3). Therefore, irrespective of the choice of γ and α , we have the approximation

$$\begin{aligned} Q_\alpha^N \simeq & \frac{1}{2} \sigma_N^{2\alpha} \left(\|\tilde{S} x^N\|_{\mathcal{C}^N}^2 - \|\tilde{S}^N (\mathcal{C}^N)^{1/2} z^N\|_{\mathcal{C}^N}^2 \right) - 2\sigma_N^{\alpha-1} \langle (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle \\ & - 2\sigma_N^{2\alpha-2} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 - \sigma_N^{2\alpha-1} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N} \\ & - \sigma_N^{3\alpha-1} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} - \frac{1}{2} \sigma_N^{4\alpha-2} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2 \\ & - \sigma_N^{3\alpha-2} \left(1 - \frac{\sigma_N^2}{2} \right) \langle \tilde{S}^N x^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N}. \end{aligned} \quad (5.6)$$

We further observe that, in stationarity (and again irrespective of the choice of γ, α and S^N) the term

$$\frac{1}{2}\sigma_N^{2\alpha} \left(\|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 - \|\tilde{S}^N (\mathcal{C}^N)^{1/2} z^N\|_{\mathcal{C}^N}^2 \right)$$

is smaller than the term

$$-2\sigma_N^{2\alpha-2} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2$$

in the sense that it is always of lower order in N . Moreover, due to the skew-symmetry of S^N , the term (5.6) is identically zero:

$$\langle \tilde{S}^N x^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} = \langle \mathcal{C}^N S^N x^N, \mathcal{C}^N S^N \mathcal{C}^N S^N x^N \rangle_{\mathcal{C}^N} = \langle \mathcal{C}^N S^N x^N, S^N \mathcal{C}^N S^N x^N \rangle = 0.$$

Therefore we can make the further approximation

$$Q_\alpha^N \simeq -2\sigma_N^{\alpha-1} \langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle - 2\sigma_N^{2\alpha-2} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 \quad (5.7)$$

$$- \sigma_N^{2\alpha-1} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N} - \frac{1}{2} \sigma_N^{4\alpha-2} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2 \quad (5.8)$$

$$- \sigma_N^{3\alpha-1} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} =: R_\alpha^N. \quad (5.9)$$

As a result of the above reasoning we then have the heuristic approximation

$$Q_\alpha^N \simeq R_\alpha^N. \quad (5.10)$$

We now heuristically argue that the only sensible choice for γ is $\gamma = 1/6$ if $\alpha \geq 2$ and $\gamma \geq 1/6$ if $\alpha < 2$. In order to do so we look at the decomposition (5.1) of Q^N and recall the definition (3.5) of the acceptance probability; we then write

$$\beta^N = 1 \wedge (e^{\bar{Q}^N} e^{Q_\alpha^N}).$$

Let us start by looking at \bar{Q}^N : if we start the chain in stationarity, i.e. $x_0^N \sim \pi^N$ then $x_k^N \sim \pi^N$ for every $k \geq 0$. In particular, if $x^N \sim \pi^N$ then x^N can be represented as $x^N = \sum_{i=1}^N \lambda_i \rho_i \varphi_i$, where ρ_i are i.i.d. $\mathcal{N}(0, 1)$. We can therefore use the law of large numbers and observe that

$$\|x^N\|_{\mathcal{C}^N}^2 = \sum_{i=1}^N |\rho_i|^2 \simeq N.$$

Similarly, by the Central Limit Theorem (CLT) the term $\langle x^N, (\mathcal{C}^N)^{1/2} z^N \rangle_{\mathcal{C}^N}$ is $O(N^{1/2})$ and converges to a standard Gaussian as $N \rightarrow \infty$. Again by the CLT we can see that also the term $(\|x^N\|_{\mathcal{C}^N}^2 - \|z^N\|^2)$ is $O(N^{1/2})$ in stationarity. Therefore, recalling (3.4), one has

$$\bar{Q}^N \simeq -\frac{\ell^3}{4N^{3\gamma}} \langle x^N, (\mathcal{C}^N)^{1/2} z^N \rangle_{\mathcal{C}^N} - \frac{\ell^6}{32N^{6\gamma}} \|x^N\|_{\mathcal{C}^N}^2.$$

With these observations in place we can then argue the following:

- If $\gamma = 1/6$ then \bar{Q}^N is $O(1)$ (in particular, for large N it converges to a Gaussian with finite mean and variance, see (5.19)); therefore $e^{\bar{Q}^N}$ is $O(1)$ as well
- If $\gamma < 1/6$ then $\bar{Q}^N \rightarrow -\infty$ (more precisely $\mathbb{E}\bar{Q}^N \rightarrow -\infty$), hence $e^{\bar{Q}^N} \rightarrow 0$
- If $\gamma > 1/6$ then $\bar{Q}^N \rightarrow 0$, hence $e^{\bar{Q}^N} \rightarrow 1$

Let us now informally discuss the acceptance probability β^N in each of the above three cases, taking into account the behaviour of Q_α^N as well. To this end it is worth noting the following:

$$\mathbb{E}_{\pi^N} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 = \mathbb{E}_{\pi^N} \|(\mathcal{C}^N)^{1/2} S^N x^N\|^2 = \mathcal{V}_S^N = \mathbb{E}_{\pi^N} |\langle \mathcal{C}^{1/2} z^N, S^N x^N \rangle|^2 \quad (5.11)$$

$$\mathbb{E}_{\pi^N} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2 = \mathcal{V}_{S\tilde{S}} = \mathbb{E}_{\pi^N} |\langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N}|^2, \quad (5.12)$$

where the first equality in (5.11) holds because

$$\|(\mathcal{C}^N)^{1/2} S^N x^N\|^2 = \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 \quad \text{for all } x,$$

and the others follow from (2.5). By (5.11)- (5.12) and (5.7) - (5.9), as $N \rightarrow \infty$, the (average, given x , of the) quantity Q_α^N can only tend to zero, go to $-\infty$ or be $O(1)$, but it cannot diverge to $+\infty$. With this in mind,

- a):** If $\gamma = 1/6$ (that is, \bar{Q}^N is $O(1)$) then one has the following two sub-cases
 - a1):** $\gamma = 1/6$ and either $Q_\alpha^N \rightarrow 0$ or Q_α^N is $O(1)$. In this case the limiting acceptance probability does not degenerate to zero (and does not tend to one)
 - a2):** $\gamma = 1/6$ and $Q_\alpha^N \rightarrow -\infty$. In this case $\beta^N \rightarrow 0$.
- b):** If $\gamma < 1/6$ then e^{Q^N} can only tend to zero (if $Q_\alpha^N \rightarrow 0$, to $-\infty$ or is $O(1)$). This is not a good regime as in this case β^N would overall tend to zero.
- c):** If $\gamma > 1/6$ we have three sub-cases
 - c1):** $\gamma > 1/6$ and Q_α^N is $O(1)$. In this case β^N is $O(1)$.
 - c2):** $\gamma > 1/6$ and $Q_\alpha^N \rightarrow 0$, in which case $\beta^N \rightarrow 1$.
 - c3):** $\gamma > 1/6$ and $Q_\alpha^N \rightarrow -\infty$, in which case β^N degenerates to zero.

We clearly want to rule out all the cases in which β^N tends to zero, (because this implies that the chain is getting stuck). Therefore, if we want to “optimize” the behaviour of the chain, the only good choices are a1), c1) and, potentially c2). However, considering that the cost of running the chain is either 2γ (when $\alpha \geq 2$) or $\alpha\gamma$ (when $\alpha < 2$), the case $\gamma > 1/6$ is always going to be sub-optimal if $\alpha \geq 2$; which is why in Theorem 4.2 we only consider $\gamma > 1/6$ if $\alpha < 2$.

We therefore want to make assumptions on S^N , γ and α to guarantee that our analysis falls only into one of the good regimes, *i.e.*, to guarantee that Q_α^N is of order one (or tends to zero). Assumption 5.1 and Assumption 5.2 below are enforced in order to guarantee that this is indeed the case.

To explain the nature of Assumption 5.1 below, we recall the approximation (5.10) and the expression for R_α^N , namely:

$$\begin{aligned} R_\alpha^N = & -2 \frac{\ell^{\alpha-1}}{N^{(\alpha-1)\gamma}} \langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle - 2 \frac{\ell^{2(\alpha-1)}}{N^{2(\alpha-1)\gamma}} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 \\ & - \frac{\ell^{2\alpha-1}}{N^{(2\alpha-1)\gamma}} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N} - \frac{1}{2} \frac{\ell^{2(2\alpha-1)}}{N^{2(2\alpha-1)\gamma}} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2 \\ & - \frac{\ell^{3\alpha-1}}{N^{(3\alpha-1)\gamma}} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} \end{aligned} \quad (5.13)$$

Roughly speaking, Assumption 5.1 requires that a CLT should hold for all the scalar products in the above expressions, Assumption 5.2 requires a Law of Large Numbers (LLN) to hold for the remaining terms. We will first state such assumptions and then comment on them in Remark 5.5.

We recall that $\tilde{S}^N = \mathcal{C}^N S^N$ so, since \mathcal{C}^N is fixed by the problem and invertible, the following assumptions can be equivalently rephrased in terms of either \tilde{S}^N or S^N .

Assumption 5.1. *The entries of the matrix S^N do not depend on N . There exist positive constants $0 \leq d_1, d_2, d_3 < \infty$ such that the sequence of matrices $\{S^N\}_N$ (and the related $\{\tilde{S}^N\}_N$) satisfies the following conditions, for some $\alpha \geq 1$:*

$$\begin{aligned} i) \quad & \frac{1}{N^{(\alpha-1)\gamma}} \langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle \xrightarrow{\mathcal{D}} \mathcal{N}(0, d_1) \\ ii) \quad & \frac{1}{N^{(2\alpha-1)\gamma}} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N} \xrightarrow{\mathcal{D}} \mathcal{N}(0, d_2) \\ iii) \quad & \frac{1}{N^{(3\alpha-1)\gamma}} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} \xrightarrow{\mathcal{D}} \mathcal{N}(0, d_3), \end{aligned}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

Assumption 5.2. *The parameter $\alpha \geq 1$ and the sequence of matrices $\{S^N\}_N$ are such that*

$$\begin{aligned} i) \quad & \lim_{N \rightarrow \infty} \mathbb{E}_{\pi^N} \left| \frac{\|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2}{N^{2(\alpha-1)\gamma}} - c_1 \right|^2 = 0 \\ ii) \quad & \lim_{N \rightarrow \infty} \mathbb{E}_{\pi^N} \left| \frac{\|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2}{N^{2(2\alpha-1)\gamma}} - c_2 \right|^2 = 0 \\ iii) \quad & \lim_{N \rightarrow \infty} \mathbb{E}_{\pi^N} \left| \frac{\|(\tilde{S}^N)^T (\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}}{N^{2(3\alpha-1)\gamma}} - c_3 \right|^2 = 0, \end{aligned} \tag{5.14}$$

for some constants $0 \leq c_1, c_2, c_3 < \infty$. We do not consider sequences $\{S^N\}$, such that (5.14) is satisfied with $c_1 = 0$ and $\alpha = 1$. That is, we consider sequences $\{S^N\}_N$ such that (5.14) is satisfied with $c_1 = 0$, only in the cases in which this happens for $\alpha > 1$, see Remark 5.5.

In view of (5.11)-(5.12), Assumption 5.2 implies the following three facts

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}_{\pi^N} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2}{N^{2(\alpha-1)\gamma}} = c_1 \tag{5.15}$$

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}_{\pi^N} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2}{N^{2(2\alpha-1)\gamma}} = c_2 \tag{5.16}$$

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\pi^N} \frac{\|(\tilde{S}^N)^T (\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}}{N^{2(3\alpha-1)\gamma}} = c_3. \tag{5.17}$$

Let us state a sufficient condition under which Assumption 5.1 holds.

Condition 5.3. *Let $z^N \sim \mathcal{N}(0, \text{Id}_N)$ and $x^N \sim \pi^N$. The sequence of matrices $\{S^N\}_N$ is such that for every $h = 1, \dots, N$ the σ -algebra generated by the random variable $z^{h,N} (S^N x^N)^h$ is independent of the σ -algebra generated by the sum*

$$\sum_{j=1}^{h-1} z^{j,N} (S^N x^N)^j.$$

In the above we stress that $(S^N x^N)$ is a vector in \mathbb{R}^N and $(S^N x^N)^h$ is the h -th component of such a vector. In order for Assumption 5.1 to be satisfied, the same should hold for $(\tilde{S}^N)^T \tilde{S}^N$ and for $(\tilde{S}^N)^T (\tilde{S}^N)^2$ as well.

Notice that, because \mathcal{C}^N is diagonal, if $(S^N)^2$ and $(S^N)^3$ satisfy Condition 5.3, then also $(\tilde{S}^N)^T \tilde{S}^N$ and $(\tilde{S}^N)^T (\tilde{S}^N)^2$ do. We give an example of a matrix satisfying all such conditions in Remark 5.5. Finally, for technical reasons (see proof of statement (ii) of Lemma B.2, we also assume the following.

Assumption 5.4. (i) Under π^N , the variables appearing in (5.15)- (5.17) have exponential moments, i.e., there exists a constant m , independent of N , such that

$$\mathbb{E}_{\pi^N} e^{\frac{\|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2}{N^{2(\alpha-1)\gamma}}}, \mathbb{E}_{\pi^N} e^{\frac{\|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2}{N^{2(2\alpha-1)\gamma}}}, \mathbb{E}_{\pi^N} e^{\frac{\|(\tilde{S}^N)^T (\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}}{N^{2(3\alpha-1)\gamma}}} < m.$$

(ii) There exists some $r > 1$ such that

$$\mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \frac{\lambda_j^6 |(S^N x^N)^j|^4}{N^{2\gamma(\alpha-1)}} \right)^r < \infty.$$

Remark 5.5. Let us clarify the meaning of Assumption 5.1, Assumption 5.2 and Condition 5.3.

- In Assumption 5.1, we do not allow the entries of S^N to depend explicitly on N because this can be achieved by simply changing scaling (i.e., values of α).
- Recall that in stationarity $x^{i,N} \sim \mathcal{N}(0, \lambda_i^2)$ and observe that without any assumption on S^N the addends in the sum

$$\langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle = \sum_{i=1}^N \lambda_i z^{i,N} (S^N x^N)^i = \sum_{i=1}^N \lambda_i z^{i,N} \sum_{j=1}^N S_{i,j}^N x^{j,N} \quad (5.18)$$

are non-independent and non-identically distributed random variables, so the Central Limit theorem does not necessarily hold and in general the behaviour of the above sum will depend on S^N . However, if we want Q_α^N to be order one, the validity of the CLT is a reasonable requirement to impose on S^N .

- Condition 5.3 allows simplifications in the calculation of the asymptotic variance for terms of the form (5.18). Indeed a generalized form of the CLT (see [8, Theorem 2.1]) implies that (under suitable assumptions) the asymptotic variance of the sum (5.18) is related to the following sum of conditional expectations:

$$\sum_k a_{k,N}^2, \quad \text{where } a_{k,N}^2 := \mathbb{E} \left(\lambda_k^2 (z^{k,N} (S^N x^N)^k)^2 \middle| \sum_{i=1}^{k-1} z^{i,N} (S^N x^N)^i \right).$$

This may not be easy to calculate in general. If Condition 5.3 holds then clearly the above simplifies and one can apply the Central Limit Theorem in its simplest form to the summation (5.18). In particular, if Condition 5.3 holds then each term is independent of the sum of the previous terms so we obtain that for large N (and, again, assuming that we are in stationarity)

$$\frac{1}{N^{(\alpha-1)\gamma}} \langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle_{\mathcal{C}^N} \sim \mathcal{N}(0, c_1).$$

Notice that we are requiring that also the structure of $(S^N)^2$ and $(S^N)^3$ allows for a similar simplification. So the main effect of Condition 5.3 is to obtain $c_i = d_i$ for $i \in \{1, 2, 3\}$. A matrix that satisfies all the requirements of Condition 5.3 is the matrix $S^N = [S_{ij}^N]_{i,j=1}^N$ of (4.7). If, for example, S^N is defined by (4.7) with $J_i = 1$ then $(S^N)^2 = -\text{Id}$ so $(S^N)^3 = -S^N$ and $(S^N)^4 = \text{Id}$.

- Roughly speaking, while Assumption 5.1 imposes the validity of a CLT, Assumption 5.2 requires that a Law of Large numbers should hold. This assumption is again made to guarantee that Q_α^N is of order one (or zero), which is the only case of interest, as we have argued (Because $Q_\alpha^N \simeq R_\alpha^N$, this is the same as assuming that R_α^N is $O(1)$.)
- We only consider values of α in the range $\alpha \geq 1$ for consistency: the sequence $\mathbb{E}_{\pi_N} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2$ is positive (and non identically zero), and the entries of S^N do not depend on N . Therefore there is no $0 \leq \alpha < 1$ such that Assumption 5.2 is satisfied for some finite constant c_1 .
- Finally, we make the obvious observation that, under Assumption 5.1, Assumption 5.2 and Condition 5.3 if the constant c_1 is assumed to be vanishing, $c_1 = 0$, then also $c_2 = c_3 = 0$. Therefore in this case Q_α^N tends to zero in L^2 .² If $c_1 = 0$ then we rule out the case $\alpha = 1$ because this would imply $\mathbb{E}_{\pi_N} \|\tilde{S}^N x^N\|_{\mathcal{C}^N} \rightarrow 0$, which could be achieved with a different rescaling. If at least one of these constants is non-zero then Q_α^N is $O(1)$.

□

To recap, we have that for N large, then

$$\bar{Q}^N \approx \mathcal{N}\left(-\frac{\ell^6}{32} \frac{\ell^6}{16}\right) \text{ if } \gamma = 1/6, \quad \bar{Q}^N \approx 0 \text{ if } \gamma > 1/6. \quad (5.19)$$

On the other hand, if Assumption 5.1 and Assumption 5.2 hold, then, from (5.7)-(5.12), we deduce after some calculus³ that asymptotically

$$Q_\alpha^N \approx \mathcal{N}(-\tilde{a}, \tilde{b}), \quad \gamma \geq 1/6, \quad (5.20)$$

where

$$\tilde{a} := 2\ell^{2(\alpha-1)}c_1 + \frac{1}{2}\ell^{2(2\alpha-1)}c_2 \quad (5.21)$$

$$\tilde{b} := 4\ell^{2(\alpha-1)}d_1 + 5\ell^{2(2\alpha-1)}d_2 + \ell^{2(3\alpha-1)}d_3. \quad (5.22)$$

From this we can heuristically deduce that if \tilde{Z} is a normally distributed random variable with

$$\tilde{Z} \sim \mathcal{N}\left(-\frac{\ell^6}{32} - \tilde{a}, \frac{\ell^6}{16} + \tilde{b}\right),$$

then, for N large one has

$$\bar{Q}^N + Q_\alpha^N = Q^N \approx \tilde{Z} \text{ if } \gamma = 1/6, \quad Q^N \approx \mathcal{N}(-\tilde{a}, \tilde{b}) \text{ if } \gamma > 1/6. \quad (5.23)$$

²This does not mean that the effect of the irreversible term is destroyed. It simply means that the acceptance probability will not feel it.

³These calculations are a bit long but straightforward and follow the lines of the calculation done in Lemma B.1, proof of point *vi*).

As explained in Remark 5.5 (see third bullet point), if also Condition 5.3 is enforced, then $c_i = d_i$ for $i \in \{1, 2, 3\}$; therefore, for N large, one finds that Q^N is approximately distributed like Q ,

$$Q^N \approx Q, \quad (5.24)$$

where

$$Q \sim \mathcal{N}\left(-\frac{\ell^6}{32} - a, \frac{\ell^6}{16} + b\right) \quad \text{if } \gamma = 1/6, \quad Q \sim \mathcal{N}(-a, b) \quad \text{if } \gamma > 1/6, \quad (5.25)$$

with

$$a := 2\ell^{2(\alpha-1)}c_1 + \frac{1}{2}\ell^{2(2\alpha-1)}c_2 \quad (5.26)$$

$$b := 4\ell^{2(\alpha-1)}c_1 + 5\ell^{2(2\alpha-1)}c_2 + \ell^{2(3\alpha-1)}c_3. \quad (5.27)$$

Before moving to the heuristic analysis of the drift coefficient, we set

$$h_{\mathbf{p}} := \mathbb{E}[(1 \wedge e^Q)], \quad (5.28)$$

with Q as in (5.25). The constant $h_{\mathbf{p}}$ represents the limiting average acceptance probability and it can be calculated by using the following lemma.

Lemma 5.6. *If G is a normally distributed random variable with $G \sim \mathcal{N}(\mu, \delta^2)$ ($\mu, \delta \neq 0$) then*

$$\mathbb{E}(1 \wedge e^G) = e^{\mu+\delta^2/2}\Phi\left(-\frac{\mu}{\delta} - \delta\right) + \Phi\left(\frac{\mu}{\delta}\right),$$

and

$$\mathbb{E}(G\mathbf{1}_{\{G < 0\}}) = e^{\mu+\delta^2/2}\Phi\left(-\frac{\mu}{\delta} - \delta\right),$$

where Φ is the CDF of a standard Gaussian.

Proof. By direct calculation. □

Therefore,

$$h_{\mathbf{p}} = \mathbb{E}(1 \wedge e^Q) = e^{\frac{b}{2}-a}\Phi\left(\frac{-\ell^6/32 - b + a}{\sqrt{(\ell^6/16) + b}}\right) + \Phi\left(\frac{-\ell^6/32 - a}{\sqrt{\ell^6/16 + b}}\right), \quad \text{if } \gamma = 1/6, \quad a, b \in \mathbb{R} \quad (5.29)$$

and

$$h_{\mathbf{p}} = e^{\frac{b}{2}-a}\Phi\left(\frac{a}{\sqrt{b}} - \sqrt{b}\right) + \Phi\left(-\frac{a}{\sqrt{b}}\right), \quad \text{if } \gamma > 1/6, \quad a, b \neq 0.$$

The above discussion makes it clear that the acceptance probability $h_{\mathbf{p}}$ depends on the constants c_i and therefore, ultimately, on the choice of the antisymmetric matrix S^N . Notice that if S^N takes the Jordan block diagonal form then $c_2 = c_3 = 0$ and $b = 2a$ resulting in limiting acceptance probability $h_{\mathbf{p}} = h_{\mathbf{p}}^J$, as in (4.12).

5.2. Heuristic analysis of the drift coefficient. In view of the discussion in Section 5.1, we comment separately on the case $\alpha > 2$, $\alpha = 2$ and $\alpha < 2$.

• Let us first assume that $\gamma = 1/6$ and suppose $\alpha > 2$. Then the approximate drift of the chain is given by (4.3)-(4.4), where we take $\zeta = 2$. We will prove that when $\alpha > 2$ the addend (4.4) is asymptotically small. This happens because Q^N and z^N are asymptotically independent; in particular we will show that the correlations between Q^N and z^N decay faster than $N^{-1/6}$, that is

$$N^{1/6} \mathbb{E}_k \left[\left(1 \wedge e^{Q^N} \right) \mathcal{C}^{1/2} z_{k+1}^N \right] \rightarrow 0$$

(the formal statement of the above heuristic is contained in B.17 and B.18. Therefore

$$\begin{aligned} d^N(x_k^N) &\simeq N^{1/3} \mathbb{E}_k \left[\left(1 \wedge e^{Q^N} \right) \left(-\frac{\ell^2}{2N^{1/3}} x_k^N + \frac{\ell^\alpha}{N^{\alpha/6}} \tilde{S}^N x_k^N \right) \right] \\ &= \mathbb{E}_k \left[\left(1 \wedge e^{Q^N} \right) \left(-\frac{\ell^2}{2} x_k^N \right) \right] + \mathbb{E}_k \left[\left(1 \wedge e^{Q^N} \right) \frac{\ell^\alpha}{N^{(\alpha-2)/6}} \tilde{S}^N x_k^N \right]. \end{aligned} \quad (5.30)$$

We now use the definition (5.28) and the heuristic approximation (5.24) and observe that if $\alpha > 2$ the second addend in (5.30) disappears in the limit; therefore the drift coefficient of the limiting diffusion is

$$d_{\mathbf{p}}(x) = -\frac{\ell^2}{2} h_{\mathbf{p}} x.$$

• If instead $\gamma = 1/6$ and $\alpha = 2$ then the second addend in (5.30) does not disappear in the limit; on the contrary, with the same reasoning as in the previous case, it will converge to $h_{\mathbf{p}} \ell^2 \tilde{S} x$. Moreover, the term in (4.4) is no longer asymptotically small and contributes to the limiting drift. The reason why this happens can be seen by recalling that $Q^N = \bar{Q}^N + Q_\alpha^N$: clearly, the correlations between \bar{Q}^N and z^N are not affected by the choice of α ; however the value of α does affect the decay of the correlations between Q_α^N and z^N .

In B.2 we present a calculation which shows that (under appropriate conditions on S^N), if $\alpha = 2$ then

$$N^{1/6} \ell \mathbb{E}_k \left[\left(1 \wedge e^{Q^N} \right) \mathcal{C}^{1/2} z_{k+1}^N \right] \simeq -2\ell^2 \nu_{\mathbf{p}} \tilde{S} x_k$$

where $\nu_{\mathbf{p}}$ is a real constant, namely

$$\nu_{\mathbf{p}} := \mathbb{E} e^{Q^N} \mathbf{1}_{\{Q < 0\}} \stackrel{\text{if } \gamma=1/6}{=} e^{\frac{b}{2}-a} \Phi \left(\frac{-\ell^6/32 - b + a}{\sqrt{(\ell^6/16) + b}} \right).^4 \quad (5.31)$$

Thus we find that the drift of the limiting diffusion is

$$d(x) = -\frac{\ell^2}{2} h_{\mathbf{p}} x + \tau_{\mathbf{p}} \ell^2 \tilde{S} x,$$

where

$$\tau_{\mathbf{p}} := h_{\mathbf{p}} - 2\nu_{\mathbf{p}}. \quad (5.32)$$

By Lemma 5.6 one can also see that, again if $\gamma = 1/6$,

$$\tau_{\mathbf{p}} = -h_{\mathbf{p}} + 2\Phi \left(\frac{-\frac{\ell^6}{32} - a}{\sqrt{\frac{\ell^6}{16} + b}} \right). \quad (5.33)$$

⁴Having used Lemma 5.6 for the last equality.

If the constant $c_1 = 0$ or if more generally $b = 2a$ (e.g., in the Jordan block diagonal form case), then $\tau_{\mathbf{p}} = 0$ and the limiting drift reduces to

$$d(x) = -\frac{\ell^2}{2} h_{\mathbf{p}} x.$$

Notice that the constant $h_{\mathbf{p}}$ can never vanish in this regime.

• As already observed, if $\gamma \geq 1/6$ and $1 \leq \alpha < 2$ we need to scale the algorithm differently; in particular, the approximate drift is given this time by

$$d_{\mathbf{p}}^N(x_k^N) \simeq N^{\alpha\gamma} \mathbb{E}_k \left[(1 \wedge e^{Q^N}) \left(-\frac{\ell^2}{2N^{2\gamma}} x_k^N + \frac{\ell^\alpha}{N^{\alpha\gamma}} \tilde{S}^N x_k^N \right) \right] + N^{\alpha\gamma} \mathbb{E}_k \left[(1 \wedge e^{Q^N}) \frac{\ell}{N^\gamma} (\mathcal{C}^N)^{1/2} z_{k+1}^N \right] \quad (5.34)$$

Therefore in this case ($\alpha < 2$) the first term in (5.34) asymptotically disappears. As for the third addend, this is again asymptotically not small (see Lemma B.2, part (ii) of B) and the second addend contributes with a term $\ell^\alpha h_{\mathbf{p}} \tilde{S}x$. Therefore, the limiting drift is

$$d(x) = \tau_{\mathbf{p}} \ell^\alpha \tilde{S}x,$$

with $\tau_{\mathbf{p}}$ defined as in (5.32). Again, if $b = 2a$ (which happens e.g. for $c_1 = c_2 = c_3 = 0$), then $\tau_{\mathbf{p}} = 0$. Therefore, in this case the limit of the chain would simply be $\dot{x} = 0$.

5.3. Heuristic analysis of the diffusion coefficient. The heuristic argument for approximate diffusion coefficient is relatively straightforward.

$$\begin{aligned} M_k^N &= N^{\zeta\gamma/2} [x_{k+1}^N - x_k^N - \mathbb{E}_k(x_{k+1}^N - x_k^N)] \\ &= N^{\zeta\gamma/2} [\tilde{\beta}^N(y_{k+1}^N - x_k^N) - \mathbb{E}_k[\tilde{\beta}^N(y_{k+1}^N - x_k^N)]] \\ &= N^{\zeta\gamma/2} \left[\tilde{\beta}^N \left(-\frac{\ell^2}{2N^{2\gamma}} x_k^N + \frac{\ell^\alpha}{2N^{\alpha\gamma}} \tilde{S}^N x_k^N + \frac{\ell}{N^\gamma} \mathcal{C}^{1/2} z_{k+1}^N \right) \right. \\ &\quad \left. - \mathbb{E}_k[\tilde{\beta}^N \left(-\frac{\ell^2}{2N^{2\gamma}} x_k^N + \frac{\ell^\alpha}{2N^{\alpha\gamma}} \tilde{S}^N x_k^N + \frac{\ell}{N^\gamma} \mathcal{C}^{1/2} z_{k+1}^N \right)] \right] \\ &= N^{(\zeta/2-2)\gamma} \left(-\frac{\ell^2}{2} \tilde{\beta}^N x_k^N \right) + N^{(\zeta/2-\alpha)\gamma} \left(\ell^\alpha \tilde{\beta}^N \tilde{S}^N x_k^N \right) + N^{(\zeta/2-1)\gamma} \left(\ell \tilde{\beta}^N \mathcal{C}^{1/2} z_{k+1}^N \right) \\ &\quad - N^{-\zeta\gamma/2} d^N(x_k^N) \end{aligned} \quad (5.35)$$

By Lemma A.2, we get that in the case $\gamma = 1/6$, $\alpha \geq 2$ and $\zeta = 2$, M_k^N behaves like $N(0, \ell^2 h_{\mathbf{p}} \mathcal{C})$; in the case $\gamma \geq 1/6$, $\alpha < 2$ and $\zeta = \alpha$ and M_k^N converges to zero.

6. ON NUMERICAL COMPUTATION OF ACCEPTANCE PROBABILITY

In this section we aim to show via concrete examples how to construct sequences of matrices S^N that satisfy the assumptions of the paper. In addition we calculate in these examples the constants d_i, c_i for $i = 1, 2, 3$ that are necessary in order to numerically approximate the acceptance probabilities in each case.

Let us first discuss how one can construct S^N so that it satisfies the assumptions of the paper, but at the same time leads to non-zero constant c_1 . Let us again consider a Jordan block-diagonal form for S^N of (4.7) but now set $S_1^N = [S_{ij}]_{i,j=1}^N$ with $S_{ij} = J_i$ and $S_{ji} = -J_i$ for $j = i + 1$ and $i = 1, 3, 5, 7, \dots, N - 1$ and $S_{ij} = 0$ otherwise with J_i to be chosen.

Let us set $\lambda_j = j^{-k}$ for $k > 1/2$. Then, we can compute

$$\begin{aligned}\mathbb{E}_{\pi_N} \|\tilde{S}_1 x^N\|_{\mathcal{C}}^2 &= \sum_{i=1}^N \lambda_i^2 \sum_{j=1}^N (S_{ij})^2 \lambda_j^2 = \sum_{i \text{ odd}} i^{-2k} |S_{i,i+1}|^2 (i+1)^{-2k} + \sum_{i \text{ even}} i^{-2k} |S_{i,i-1}|^2 (i-1)^{-2k} \\ &= 2 \sum_{i=1}^N (2i-1)^{-2k} J_i^2 (2i)^{-2k}\end{aligned}$$

In order to make sure that $c_1 \neq 0$, let us choose $J_i = (2i-1)^k (2i)^k i^{[2(\alpha-1)\gamma-1]/2}$ with $\alpha > 1$. Then, we have that $\mathbb{E}_{\pi_N} \|\tilde{S}_1 x^N\|_{\mathcal{C}}^2 = 2 \sum_{i=1}^N i^{2(\alpha-1)\gamma-1}$. We obtain that

$$c_1 = \lim_{N \rightarrow \infty} \mathbb{E}_{\pi_N} \frac{\|\tilde{S}_1 x^N\|_{\mathcal{C}^N}^2}{N^{2(\alpha-1)\gamma}} = \lim_{N \rightarrow \infty} \frac{1}{N} \frac{2 \sum_{i=1}^N \left(\frac{i}{N}\right)^{2(\alpha-1)\gamma-1}}{N^{2(\alpha-1)\gamma}} = 2 \int_0^1 x^{2(\alpha-1)\gamma-1} dx = \frac{1}{(\alpha-1)\gamma}.$$

Next we compute c_2 for this particular choice of the S matrix. We have

$$\begin{aligned}\mathbb{E}_{\pi_N} \|\tilde{S}_1^2 x^N\|_{\mathcal{C}^N}^2 &= \sum_{i=1}^N \lambda_i^2 \sum_{j=1}^N ((S\tilde{S})_{ij})^2 \lambda_j^2 = \sum_{i=1}^N \lambda_i^2 ((S\tilde{S})_{ii})^2 \lambda_i^2 = 2 \sum_{i \text{ odd}} \lambda_i^4 \lambda_{i+1}^4 J_{(i+1)/2}^4 \\ &= 2 \sum_{i \text{ odd}} i^{-4k} (i+1)^{-4k} (2(i+1)/2-1)^{4k} (2(i+1)/2)^{4k} \left(\frac{i+1}{2}\right)^{2[2(\alpha-1)\gamma-1]} \\ &= 2 \sum_{i \text{ odd}} i^{-4k} (i+1)^{-4k} (i)^{4k} (i+1)^{4k} \left(\frac{i+1}{2}\right)^{2[2(\alpha-1)\gamma-1]} \\ &= 2 \sum_{i \text{ odd}} \left(\frac{i+1}{2}\right)^{2[2(\alpha-1)\gamma-1]} \\ &= 2 \sum_{i=1}^N i^{2[2(\alpha-1)\gamma-1]}\end{aligned}$$

Similarly to the computation for c_1 , we obtain that $c_2 = 0$. As far as c_3 is concerned, we obtain by a computation similar to the computation for c_2 that

$$c_3 = \lim_{N \rightarrow \infty} \mathbb{E}_{\pi_N} \frac{\|\tilde{S}_1^T \tilde{S}_3^2 x\|_{\mathcal{C}}}{N^{2(3\alpha-1)\gamma}} = 0$$

For the same reasons as in Remark 5.5 we also have $d_i = c_i$ for $i = 1, 2, 3$. Hence, we have obtained that in this example $d_1 = c_1 = \frac{1}{(\alpha-1)\gamma}$ and $d_2 = c_2 = d_3 = c_3 = 0$. Then, this implies that $Q \sim \mathcal{N}\left(-\frac{\ell^6}{32} - a, \frac{\ell^6}{16} + b\right)$ in (5.25), with $a = 2\ell^{2(\alpha-1)} \frac{1}{(\alpha-1)\gamma}$ and $b = 4\ell^{2(\alpha-1)} \frac{1}{(\alpha-1)\gamma}$. Hence, in this case $b = 2a$ and thus $\tau_{\mathbf{p}} = 0$.

Next we discuss optimal acceptance probabilities. Since in this case $b = 2a$ which then implies $\tau_{\mathbf{p}} = 0$, we can answer this question in the case of regimes i), i.e., $\alpha > 2$ and ii) $\alpha = 2$. In these cases we can find the ℓ that maximizes the speed of the limiting diffusion, i.e., $\ell^2 h_{\mathbf{p}}$ and then for the maximizing ℓ compute the limiting acceptance probability $h_{\mathbf{p}}$ as given by (5.29). In Table 3 we record optimal acceptance probabilities $h_{\mathbf{p}}$ derived in (5.29) for different choices for α when $\gamma = 1/6$ in the case the matrix S being S_1 as defined above.

α	2	4	6	8	10	15	30
optimal $h_{\mathbf{p}}$	0.234	0.574	0.702	0.767	0.803	0.848	0.884

TABLE 3. Optimal acceptance probability when $S^N = S_1^N$ for different α 's when $\gamma = 1/6$.

In particular, Table 3 shows that by increasing α , we can increase the optimal acceptance probability as desired.

Next, we discuss two cases where $c_1 = c_2 = c_3 = 0$. In particular, one can define the matrix $S_2^N = [S_{ij}]_{i,j=1}^N$ such that $S_{ij} = 1$ and $S_{ji} = -1$ for $j = i + 1$ and $i = 1, 2, 3, 4, \dots, N$ and $S_{ij} = 0$ otherwise. A similar alternative construction is to set $S_3^N = [S_{ij}]_{i,j=1}^N$ with $S_{ij} = 1$ and $S_{ji} = -1$ for $j = i + 1$ and $i = 1, 3, 5, \dots, N - 1$ and $S_{ij} = 0$ otherwise. In both cases S_2^N and S_3^N one can check that $c_i = d_i = 0$ for all $i = 1, 2, 3$. This then implies that $Q \sim \mathcal{N}\left(-\frac{\ell^6}{32}, \frac{\ell^6}{16}\right)$ in (5.25), as $a = b = 0$ in this case. Hence, one recovers that in this case the optimal acceptance probability for both choices S_1^N and S_2^N is achieved at 0.574 as in the classical work [26].

In Section 7 we present simulation studies based on using S_1^N , S_2^N and S_3^N confirming and illustrating the theoretical results.

7. SIMULATION STUDIES

The goal of this section is to illustrate the theoretical findings of this paper via simulation studies. Upon close inspection of Theorem 4.2 it becomes apparent that closed form expressions of quantities such as optimal acceptance probabilities and optimal scalings are hard to obtain and dependent on the choice of the sequence S^N . For this reason we will resort to simulations.

As a measure of performance we choose similarly to [26, 27] the expected squared jumping distance (ESJD) of the algorithm defined as

$$\text{ESJD} = \mathbb{E} |X_{k+1}^1 - X_k^1|^2$$

It is easy to see that maximizing ESJD is equivalent to minimizing the first order lag autocorrelation function, $\rho_1 = \text{Corr}_\pi(X_{k+1}, X_k)$. Hence, motivated by [27] and based on Theorem 4.2 we define the computational time as

$$CT_{\mathbf{p}} = -\frac{1}{N^{\zeta\gamma} \log \rho_1} \quad (7.1)$$

where $\zeta = 2$ if $\alpha \geq 2$ and $\zeta = \alpha$ if $\alpha < 2$. We seek to minimize $CT_{\mathbf{p}}$. A word of caution is necessary here. Even though $CT_{\mathbf{p}}$ as a measure of performance is easily justifiable in the case that one has a diffusion limit, *i.e.*, when $\alpha \geq 2$, the situation is more complex for $\alpha < 2$. By Theorem 4.2 if $\alpha < 2$, the algorithm converges to a deterministic ODE limit and not to a diffusion limit. Hence, in this case the absolute jumping distance $|X_{k+1}^1 - X_k^1|$ may make more sense as a measure of performance in the limit as $N \rightarrow \infty$. However, in the deterministic case, maximizing the absolute jumping distance or the squared jumping distance are equivalent tasks. Hence, working with the acceptance probability and with ESJD and as a consequence with $CT_{\mathbf{p}}$ as defined in (7.1) may be reasonable criterion on quantifying convergence.

The simulations presented below were performed using a MPI C parallel code. The numerical results that follow report sample averages and sample standard deviations of quantities such as acceptance probability and different observables. The reported values were obtained based on 1600 independent repetitions of MALA and ipMALA algorithm with 10^4 steps for each independent realization. For each independent realization of the algorithm the numerical values selected are the ones corresponding to minimizing $CT_{\mathbf{p}}$. We also remark here that in all of the simulation studies $CT_{\mathbf{p}}$ was a convex function of the acceptance probability (this resembles the behavior in the standard MALA case, see [27]).

Below we report a statistical estimator and its corresponding empirical standard deviation for the optimal acceptance probability. In addition, to demonstrate convergence to the equilibrium we also computed a few observables. We report below statistical estimators along with their corresponding statistical standard deviations for

$$\theta_2 = \mathbb{E}_{\pi^N} \|X\|^2, \text{ and } \theta_3 = \mathbb{E}_{\pi^N} \sum_{i=1}^N X_i^3,$$

where with some abuse of notation we have denoted by X_i the i^{th} component of the X vector. In regards to the measure $\pi^N \sim N(0, \mathcal{C}^N)$, where $\mathcal{C}^N = \text{diag}(\lambda_1^2, \dots, \lambda_N^2)$ with $\lambda_i^2 = i^{-2}$.

Let us first study the behavior of the algorithm when $S^N = S_1^N$, as defined in Section 6. In this case $c_1 = \frac{1}{(\alpha-1)\gamma} \neq 0$ and $c_2 = c_3 = 0$. However, as we saw in Table 3 the effect of the value of α on the optimal acceptance probabilities is significant.

Let us demonstrate the behavior via a number of simulation studies reported in Tables 4-6. As we described in Section 6, such a computation makes sense mainly in the case $\alpha \geq 2$ and $\gamma = 1/6$. Of course one can compute via simulation empirical values for any value of $\alpha > 1$ and $\gamma > 0$. The acceptance probabilities in the case $\alpha \in (1, 2)$ were very small, leading to estimators with large variances. Hence, we only report values for larger values of α where one may expect to have results that are comparable to what standard MALA gives. For completeness, we also present data for $\gamma = 1/2$, even though for standard MALA the optimal choice is $\gamma = 1/6$.

Notice that the estimates for acceptance probabilities $\hat{h}_{\mathbf{p}}$ align very well with the theoretical predictions that appear in Table 3. In particular, as α increases the optimal acceptance probability also increases, in accordance to what the theory predicts.

To visualize the situation, in Figure 1 we have plotted α versus the empirical acceptance probabilities $\hat{h}_{\mathbf{p}}$ for $\gamma = 1/6$. As expected by the theory, as α increases, $\hat{h}_{\mathbf{p}}$ increases.

In Figure 2 we have plotted α versus the statistic $\widehat{CT}_{\mathbf{p}}$. In each case, standard MALA is represented by the rightmost value for α . As expected by the theory, in each of the cases, as α increases $\widehat{CT}_{\mathbf{p}}$ converges to the corresponding value of standard MALA.

Next we record simulations for the antisymmetric matrix $S^N = S_2^N$, defined in Section 6. We remark here that we also did the same simulations with the $S = S_3^N$ and the numerical results were statistically the same. So, we only report the ones for $S^N = S_2^N$, see Tables 7-9.

In the case of $S^N = S_2^N$, we have that $c_i = d_i = 0$ and thus the optimal acceptance probability in the limit as $N \rightarrow \infty$ is the same as with standard MALA, i.e. around 0.574. However, as the Tables 7-9 demonstrate when $\gamma = 1/6$ (the optimal choice for standard MALA) this is being realized for much larger values of N for ipMALA compared to MALA. On the other hand if we take $\gamma > 1/6$ (which is not optimal for standard MALA), e.g.,

$\gamma = 1/6$	$(\widehat{h}_{\mathbf{p}}, \widehat{sd}(h_{\mathbf{p}}))$	$(\widehat{\theta}_2, \widehat{sd}(\theta_2))$	$(\widehat{\theta}_3, \widehat{sd}(\theta_3))$	$\widehat{CT}_{\mathbf{p}}$
standard MALA	(0.541, 0.043)	(1.073, 0.022)	(-0.001, 0.061)	0.725
$\alpha = 2$	(0.229, 0.056)	(0.935, 0.081)	(-0.001, 0.211)	8.139
$\alpha = 4$	(0.566, 0.065)	(1.026, 0.034)	(0.006, 0.141)	3.367
$\alpha = 6$	(0.686, 0.063)	(1.045, 0.072)	(0.003, 0.101)	1.946
$\alpha = 8$	(0.751, 0.051)	(1.051, 0.026)	(0.0001, 0.101)	1.418
$\alpha = 10$	(0.784, 0.052)	(1.056, 0.061)	(0.004, 0.172)	1.151
$\alpha = 15$	(0.821, 0.045)	(1.059, 0.031)	(-0.002, 0.096)	0.951
$\alpha = 30$	(0.855, 0.026)	(1.069, 0.025)	(-0.003, 0.087)	0.798
$\gamma = 1/2$				
standard MALA	(0.753, 0.008)	(1.072, 0.020)	(-0.002, 0.068)	0.198
$\alpha = 2$	(0.205, 0.066)	(0.869, 0.118)	(0.0002, 0.339)	9.419
$\alpha = 4$	(0.555, 0.092)	(0.979, 0.087)	(-0.012, 0.242)	2.022
$\alpha = 6$	(0.691, 0.083)	(0.997, 0.056)	(0.002, 0.174)	1.411
$\alpha = 8$	(0.761, 0.068)	(1.001, 0.038)	(-0.008, 0.166)	1.232
$\alpha = 10$	(0.808, 0.064)	(1.015, 0.036)	(0.002, 0.168)	1.051
$\alpha = 15$	(0.866, 0.054)	(1.023, 0.032)	(-0.0003, 0.156)	0.551
$\alpha = 30$	(0.924, 0.046)	(1.038, 0.067)	(-0.002, 0.206)	0.213

TABLE 4. Dimension $N = 10$ and $S^N = S_1^N$.

$\gamma = 1/6$	$(\widehat{h}_{\mathbf{p}}, \widehat{sd}(h_{\mathbf{p}}))$	$(\widehat{\theta}_2, \widehat{sd}(\theta_2))$	$(\widehat{\theta}_3, \widehat{sd}(\theta_3))$	$\widehat{CT}_{\mathbf{p}}$
standard MALA	(0.538, 0.052)	(1.054, 0.025)	(0.004, 0.089)	0.912
$\alpha = 2$	(0.216, 0.077)	(0.882, 0.177)	(-0.035, 0.451)	8.453
$\alpha = 4$	(0.563, 0.074)	(1.004, 0.041)	(0.004, 0.180)	3.143
$\alpha = 6$	(0.689, 0.068)	(1.026, 0.035)	(-0.009, 0.141)	2.024
$\alpha = 8$	(0.748, 0.064)	(1.036, 0.042)	(0.003, 0.157)	1.512
$\alpha = 10$	(0.784, 0.063)	(1.048, 0.098)	(0.006, 0.181)	1.269
$\alpha = 15$	(0.824, 0.045)	(1.048, 0.031)	(-0.005, 0.123)	0.991
$\alpha = 30$	(0.849, 0.035)	(1.054, 0.025)	(-0.0002, 0.114)	0.843
$\gamma = 1/2$				
standard MALA	(0.849, 0.002)	(1.038, 0.037)	(0.0004, 0.147)	0.211
$\alpha = 2$	(0.222, 0.102)	(0.681, 0.111)	(0.025, 0.437)	3.396
$\alpha = 4$	(0.564, 0.013)	(0.837, 0.107)	(0.008, 0.378)	1.696
$\alpha = 6$	(0.699, 0.023)	(0.887, 0.139)	(0.003, 0.272)	1.405
$\alpha = 8$	(0.774, 0.071)	(1.001, 0.019)	(0.005, 0.267)	1.269
$\alpha = 10$	(0.816, 0.034)	(0.967, 0.056)	(-0.002, 0.263)	1.101
$\alpha = 15$	(0.873, 0.088)	(0.948, 0.101)	(0.008, 0.264)	0.781
$\alpha = 30$	(0.932, 0.073)	(0.958, 0.071)	(-0.001, 0.114)	0.583

TABLE 5. Dimension $N = 50$ and $S^N = S_1^N$.

$\gamma = 1/2$ in Table 9, and focus on the fluid regime, regime iii), we notice that $CT_{\mathbf{p}}$ is estimated to be smaller with ipMALA compared to MALA, as predicted by our theory.

To visualize the situation, in Figure 3 we have plotted α versus the statistic $\widehat{CT}_{\mathbf{p}}$. In each case, standard MALA is represented by the rightmost value for α . As we observe from Tables 7-9 and from Figure 3 and also indicated by the theory, when $\gamma > 1/6$ (in particular $\gamma = 1/2$ here), $\widehat{CT}_{\mathbf{p}}$ is smaller when $1 < \alpha < 2$. In particular, notice that for $\gamma = 1/2 > 1/6$, $\widehat{CT}_{\mathbf{p}}$ seems to obtain a minimum value around $\alpha \approx 1.8 < 2$, which is in the fluid regime.

The numerical studies presented above illustrate the theoretical findings of the paper and also show that the choice of the irreversible perturbations S can have serious consequences on the performance of the algorithm.

$\gamma = 1/6$	$(\widehat{h}_{\mathbf{p}}, \widehat{sd}(\widehat{h}_{\mathbf{p}}))$	$(\widehat{\theta}_2, \widehat{sd}(\widehat{\theta}_2))$	$(\widehat{\theta}_3, \widehat{sd}(\widehat{\theta}_3))$	$\widehat{CT}_{\mathbf{p}}$
standard MALA	(0.538, 0.057)	(1.047, 0.036)	(0.001, 0.093)	0.938
$\alpha = 2$	(0.225, 0.071)	(0.891, 0.178)	(0.003, 0.352)	7.654
$\alpha = 4$	(0.559, 0.083)	(1.003, 0.039)	(-0.008, 0.180)	3.467
$\alpha = 6$	(0.688, 0.073)	(1.019, 0.047)	(-0.001, 0.168)	2.012
$\alpha = 8$	(0.749, 0.068)	(1.032, 0.036)	(-0.006, 0.155)	1.488
$\alpha = 10$	(0.784, 0.059)	(1.034, 0.038)	(-0.001, 0.144)	1.245
$\alpha = 15$	(0.823, 0.051)	(1.043, 0.038)	(-0.008, 0.131)	0.981
$\alpha = 30$	(0.851, 0.034)	(1.053, 0.033)	(-0.0016, 0.124)	0.942
$\gamma = 1/2$				
standard MALA	(0.948, 0.002)	(1.019, 0.045)	(0.0001, 0.183)	0.211
$\alpha = 2$	(0.221, 0.101)	(0.782, 0.113)	(0.015, 0.357)	3.145
$\alpha = 4$	(0.562, 0.015)	(0.843, 0.139)	(0.012, 0.313)	1.551
$\alpha = 6$	(0.696, 0.034)	(0.841, 0.121)	(0.001, 0.275)	1.299
$\alpha = 8$	(0.771, 0.082)	(0.951, 0.082)	(-0.001, 0.217)	1.187
$\alpha = 10$	(0.817, 0.067)	(0.989, 0.063)	(-0.003, 0.311)	1.127
$\alpha = 15$	(0.873, 0.093)	(1.003, 0.102)	(-0.002, 0.235)	0.967
$\alpha = 30$	(0.937, 0.075)	(0.908, 0.101)	(-0.013, 0.111)	0.486

TABLE 6. Dimension $N = 100$ and $S^N = S_1^N$.

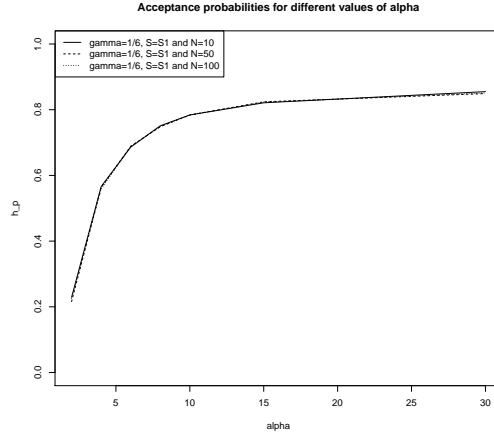


FIGURE 1. Empirical acceptance probability $\widehat{h}_{\mathbf{p}}$ versus values of α for $S^N = S_1^N$ and $\gamma = 1/6$. For reference the limiting optimal acceptance probability for standard MALA is 0.574.

In particular, in the case of $S = S_1^N$

- the theoretical acceptance probability for ipMALA is increasing as a function of α . This means that by increasing α one can increase the optimal acceptance probability considerably and combined with the fact that (a) the limit of the interpolation of the chain is a diffusion as in the standard MALA, and that (b) for large values of α the values of $\widehat{CT}_{\mathbf{p}}$ are the same for ipMALA and standard MALA, we have an algorithm with the same cost but much higher acceptance probabilities.
- the statistic $\widehat{CT}_{\mathbf{p}}$ for ipMALA converges to that for the standard MALA case as α increases. However, when $1 < \alpha < 2$ the performance seems to be worse than standard MALA due to very low optimal acceptance probabilities in this case.

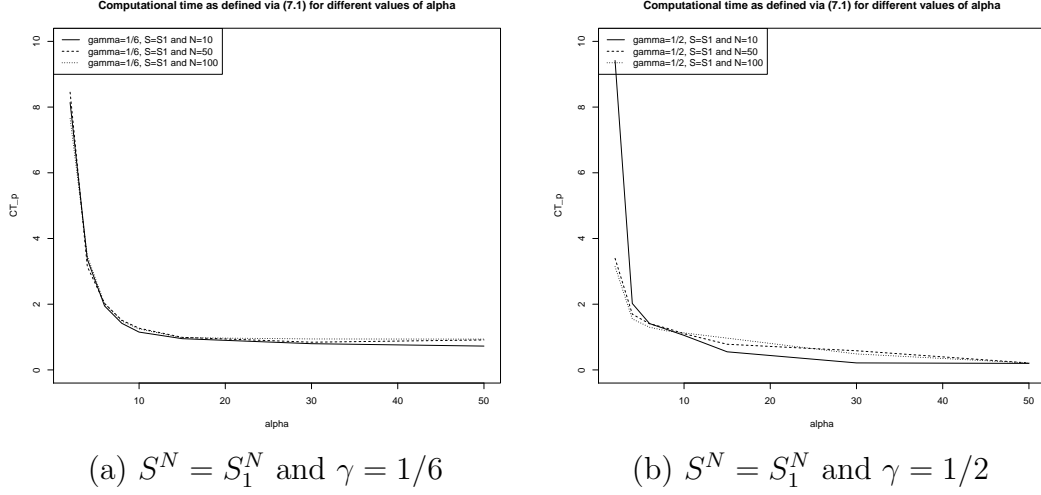


FIGURE 2. Measure $\widehat{CT}_{\mathbf{p}}$ versus values of α . In each of the figures the right-most value for α corresponds to standard MALA

$\gamma = 1/6$	$(\widehat{h}_{\mathbf{p}}, \widehat{sd}(h_{\mathbf{p}}))$	$(\widehat{\theta}_2, \widehat{sd}(\theta_2))$	$(\widehat{\theta}_3, \widehat{sd}(\theta_3))$	$\widehat{CT}_{\mathbf{p}}$
standard MALA	(0.537, 0.043)	(1.071, 0.022)	(-0.0002, 0.062)	0.725
$\alpha = 1.2$	(0.464, 0.039)	(1.064, 0.025)	(0.004, 0.071)	1.289
$\alpha = 1.5$	(0.472, 0.041)	(1.063, 0.025)	(-0.0016, 0.073)	1.175
$\alpha = 1.8$	(0.481, 0.041)	(1.061, 0.025)	(-0.0015, 0.074)	1.070
$\alpha = 2.0$	(0.488, 0.040)	(1.061, 0.025)	(0.0005, 0.077)	1.005
$\alpha = 2.5$	(0.507, 0.043)	(1.059, 0.024)	(-0.0041, 0.076)	1.038
$\alpha = 3.0$	(0.535, 0.044)	(1.061, 0.023)	(0.00068, 0.077)	1.063
$\gamma = 1/2$				
standard MALA	(0.753, 0.008)	(1.071, 0.020)	(-0.0006, 0.066)	0.427
$\alpha = 1.2$	(0.646, 0.008)	(1.065, 0.024)	(-0.0025, 0.082)	0.663
$\alpha = 1.5$	(0.649, 0.008)	(1.066, 0.024)	(0.0025, 0.082)	0.466
$\alpha = 1.8$	(0.653, 0.009)	(1.063, 0.024)	(-0.0011, 0.081)	0.327
$\alpha = 2.0$	(0.655, 0.009)	(1.064, 0.024)	(0.0010, 0.083)	0.556
$\alpha = 2.5$	(0.662, 0.011)	(1.063, 0.023)	(0.0011, 0.084)	0.549
$\alpha = 3.0$	(0.668, 0.012)	(1.064, 0.023)	(0.00317, 0.078)	0.541

TABLE 7. Dimension $N = 10$ and $S^N = S_2^N$.

- for large N , the estimated standard deviation of the statistics that were computed, θ_2 and θ_3 , appear to be slightly larger than the corresponding estimated standard deviation of the statistics for the standard MALA case, but certainly of the same order.

On the other hand, in the case of $S = S_2^N$

- the theoretical acceptance probability for ipMALA is the same as that of standard MALA, i.e., around 0.574, even though this seems to be realizable in practice in very high dimensions as even for $N = 100$, the estimated optimal acceptance probabilities were lower.
- the statistic $\widehat{CT}_{\mathbf{p}}$ has lower values for $\gamma = 1/2 > 1/6$ and $1 < \alpha < 2$ when compared to the corresponding values for the standard MALA case. This can be explained by

$\gamma = 1/6$	$(\widehat{h}_{\mathbf{p}}, \widehat{sd}(h_{\mathbf{p}}))$	$(\widehat{\theta}_2, \widehat{sd}(\theta_2))$	$(\widehat{\theta}_3, \widehat{sd}(\theta_3))$	$\widehat{CT}_{\mathbf{p}}$
standard MALA	(0.538, 0.052)	(1.054, 0.025)	(−0.0035, 0.086)	0.900
$\alpha = 1.2$	(0.463, 0.047)	(1.046, 0.028)	(−0.0023, 0.100)	1.826
$\alpha = 1.5$	(0.476, 0.048)	(1.046, 0.027)	(−0.0021, 0.097)	1.485
$\alpha = 1.8$	(0.489, 0.049)	(1.046, 0.027)	(0.0016, 0.096)	1.209
$\alpha = 2.0$	(0.496, 0.052)	(1.046, 0.031)	(0.0017, 0.094)	1.053
$\alpha = 2.5$	(0.518, 0.051)	(1.046, 0.026)	(−0.0028, 0.096)	1.037
$\alpha = 3.0$	(0.528, 0.053)	(1.047, 0.028)	(0.00031, 0.093)	1.021
$\gamma = 1/2$				
standard MALA	(0.949, 0.003)	(1.036, 0.038)	(−0.0015, 0.144)	0.776
$\alpha = 1.2$	(0.788, 0.006)	(1.021, 0.043)	(−0.0007, 0.169)	1.326
$\alpha = 1.5$	(0.832, 0.005)	(1.023, 0.041)	(0.0008, 0.166)	0.684
$\alpha = 1.8$	(0.865, 0.005)	(1.028, 0.041)	(−0.0043, 0.158)	0.361
$\alpha = 2.0$	(0.883, 0.005)	(1.030, 0.041)	(−0.0021, 0.158)	0.873
$\alpha = 2.5$	(0.914, 0.004)	(1.031, 0.038)	(−0.0001, 0.153)	0.828
$\alpha = 3.0$	(0.931, 0.003)	(1.034, 0.037)	(0.0002, 0.151)	0.803

TABLE 8. Dimension $N = 50$ and $S^N = S_2^N$.

$\gamma = 1/6$	$(\widehat{h}_{\mathbf{p}}, \widehat{sd}(h_{\mathbf{p}}))$	$(\widehat{\theta}_2, \widehat{sd}(\theta_2))$	$(\widehat{\theta}_3, \widehat{sd}(\theta_3))$	$\widehat{CT}_{\mathbf{p}}$
standard MALA	(0.549, 0.054)	(1.045, 0.027)	(0.0015, 0.102)	0.955
$\alpha = 1.2$	(0.473, 0.049)	(1.037, 0.029)	(−0.0012, 0.115)	2.080
$\alpha = 1.5$	(0.484, 0.054)	(1.036, 0.034)	(−0.0028, 0.111)	1.623
$\alpha = 1.8$	(0.498, 0.054)	(1.037, 0.036)	(−0.0021, 0.109)	1.268
$\alpha = 2.0$	(0.507, 0.056)	(1.041, 0.084)	(0.0074, 0.235)	1.076
$\alpha = 2.5$	(0.523, 0.054)	(1.041, 0.027)	(−0.0018, 0.106)	1.052
$\alpha = 3.0$	(0.533, 0.057)	(1.039, 0.034)	(0.00019, 0.104)	1.031
$\gamma = 1/2$				
standard MALA	(0.975, 0.002)	(1.006, 0.050)	(0.0103, 0.206)	0.970
$\alpha = 1.2$	(0.809, 0.006)	(0.981, 0.055)	(0.0039, 0.228)	1.705
$\alpha = 1.5$	(0.865, 0.005)	(0.989, 0.053)	(0.0138, 0.216)	0.785
$\alpha = 1.8$	(0.904, 0.004)	(0.993, 0.054)	(−0.0066, 0.213)	0.371
$\alpha = 2.0$	(0.922, 0.004)	(0.997, 0.049)	(0.0004, 0.215)	1.055
$\alpha = 2.5$	(0.952, 0.003)	(1.002, 0.049)	(0.0018, 0.211)	1.009
$\alpha = 3.0$	(0.966, 0.002)	(1.002, 0.049)	(−0.0016, 0.204)	0.985

TABLE 9. Dimension $N = 100$ and $S^N = S_2^N$.

the cost of the algorithm which in that case is $N^{\alpha\gamma}$ with $\alpha\gamma < 1/3$ as opposed to the $N^{1/3}$ which is the cost of the standard MALA case. As α increases, the value of $\widehat{CT}_{\mathbf{p}}$ seems to approach that of standard MALA.

- the estimated standard deviation of the statistics that were computed, θ_2 and θ_3 , are either the same or slightly larger than the corresponding estimated standard deviation of the statistics for the standard MALA case, but again of the same order.

8. ACKNOWLEDGMENTS

K.S. was partially supported by NSF CAREER AWARD nsf-dms 1550918.

APPENDIX A. PROOF OF THEOREM 4.2

In this section we present the proof of our main results. The proof is based on diffusion approximation techniques analogous to those used in [18]. In [18] the authors consider the

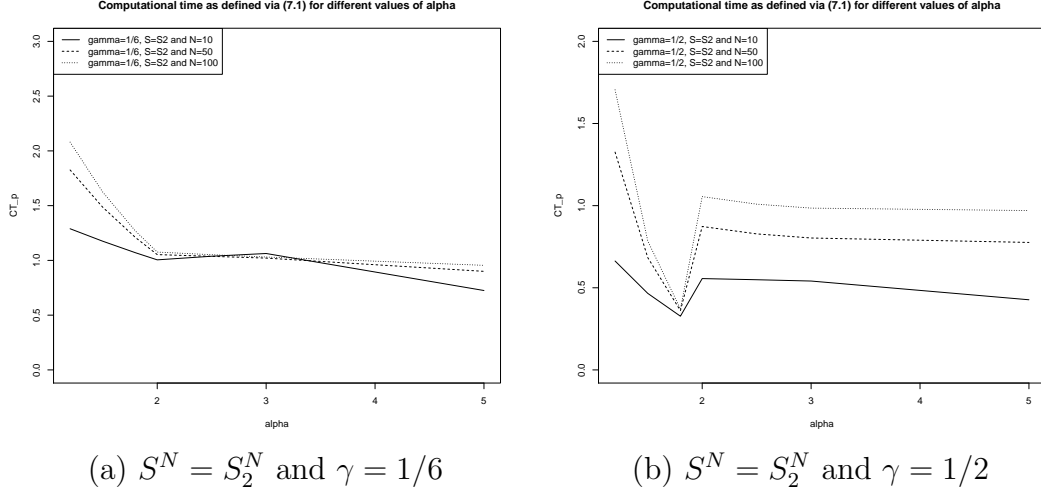


FIGURE 3. Measure $\widehat{CT}_{\mathbf{p}}$ versus values of α . In each of the figures the right-most value for α corresponds to standard MALA

MALA algorithm with reversible proposal. That is, if we fix $S = 0$ in our paper and $\Psi = 0$ in their paper, the algorithms we consider coincide. For this reason we try to adopt a notation as similar as possible to the one used in [18] and, for the sake of brevity, we detail only the parts of the proof that differ from the work [18] and just sketch the rest.

We start by recalling that, by Fernique's theorem,

$$\mathbb{E}_{\pi^N} \|x^N\|^p = \mathbb{E} \|(\mathcal{C}^N)^{1/2} z^N\|^p \lesssim 1, \quad \text{for all } p \geq 1. \quad (\text{A.1})$$

This fact will be often implicitly used without mention in the remainder of the paper.

We also recall that the chain $\{x_k^N\}_k$ that we consider is defined in (3.6); the drift-martingale decomposition of such a chain is given in equation (4.1). Let us start by recalling the definition of the continuous interpolant of the chain, equations (4.5)-(4.6), and by introducing the piecewise constant interpolant of the chain x_k^N , that is,

$$\bar{x}^{(N)}(t) = x_k^N \quad t_k \leq t < t_{k+1},$$

where $t_k = k/N^{\zeta\gamma}$. It is easy to see (see e.g. [21, Appendix A]) that

$$x^{(N)}(t) = x_0^N + \int_0^t d_{\mathbf{p}}^N(\bar{x}^{(N)}(v)) dv + w_{\mathbf{p}}^N(t), \quad (\text{A.2})$$

where

$$w_{\mathbf{p}}^N(t) := \frac{1}{N^{\zeta\gamma/2}} \sum_{j=1}^{k-1} M_j^N + N^{\zeta\gamma/2} (t - t_k) M_k^N.$$

For any $t \in [0, T]$, we set

$$\begin{aligned}\hat{w}_{\mathbf{p}}^N(t) &:= \int_0^t [d_{\mathbf{p}}^N(\bar{x}^{(N)}(v)) - d_{\mathbf{p}}(x^{(N)}(v))] dv + w_{\mathbf{p}}^N(t) \\ &= \int_0^t [d_{\mathbf{p}}^N(\bar{x}^{(N)}(v)) - d_{\mathbf{p}}(\bar{x}^{(N)}(v))] dv \\ &\quad + \int_0^t [d_{\mathbf{p}}(\bar{x}^{(N)}(v)) - d_{\mathbf{p}}(x^{(N)}(v))] dv + w_{\mathbf{p}}^N(t).\end{aligned}\tag{A.3}$$

With the above notation, we can then rewrite (A.2) as

$$x^{(N)}(t) = x_0^N + \int_0^t d_{\mathbf{p}}(x^{(N)}(v)) dv + \hat{w}_{\mathbf{p}}^N(t).\tag{A.4}$$

Let now $C([0, T]; \mathcal{H})$ denote the space of \mathcal{H} -valued continuous functions, endowed with the uniform topology and consider the map

$$\begin{aligned}\mathcal{I} : \mathcal{H} \times C([0, T]; \mathcal{H}) &\longrightarrow C([0, T]; \mathcal{H}) \\ (x_0, \eta(t)) &\longrightarrow x(t).\end{aligned}$$

That is, \mathcal{I} is the map that to every $(x_0, \eta(t)) \in \mathcal{H} \times C([0, T]; \mathcal{H})$ associates the (unique solution) of the equation

$$x(t) = x_0 + \int_0^t d_{\mathbf{p}}(x(s)) ds + \eta(t).\tag{A.5}$$

From (A.4) it is clear that $x^{(N)} = \mathcal{I}(x_0^N, \hat{w}_{\mathbf{p}}^N)$. Notice that, under our continuity assumption on \tilde{S} , \mathcal{I} is a continuous map. Therefore, in order to prove that $x^{(N)}(t)$ converges weakly to $x(t)$ (where $x(t)$ is the solution of equation (A.5) with $\eta(t) = D_{\mathbf{p}}W^{\mathcal{C}}(t)$), by the continuous mapping theorem we only need to prove that $\hat{w}_{\mathbf{p}}^N$ converges weakly to $D_{\mathbf{p}}W^{\mathcal{C}}(t)$, where $W^{\mathcal{C}}(t)$ is a \mathcal{H} -valued \mathcal{C} -Brownian motion. The weak convergence of $\hat{w}_{\mathbf{p}}^N$ to $D_{\mathbf{p}}W^{\mathcal{C}}(t)$ is a consequence of (A.3) and of Lemmata A.1 and A.2. Then, we get the statement of Theorem 4.2. The proof of Lemma A.1 and Lemma A.2 is contained in the remained of this Appendix.

Lemma A.1. *Under Assumption 5.1, Assumption 5.2 and Condition 5.3, the following holds*

$$\mathbb{E}_{\pi^N} \int_0^T \|d_{\mathbf{p}}^N(\bar{x}^{(N)}(t)) - d_{\mathbf{p}}(\bar{x}^{(N)}(t))\| dt \longrightarrow 0, \quad \text{as } N \rightarrow \infty\tag{A.6}$$

and

$$\mathbb{E}_{\pi^N} \int_0^T \|d_{\mathbf{p}}(\bar{x}^N(t)) - d_{\mathbf{p}}(x^{(N)}(t))\| dt \longrightarrow 0, \quad \text{as } N \rightarrow \infty,\tag{A.7}$$

where the function $d_{\mathbf{p}}(x)$ has been defined in the statement of Theorem 4.2.

Set now

$$\epsilon_{\mathbf{p}}^N(x) := \mathbb{E}_x \left[\left(1 \wedge e^{Q^N} \right) \mathcal{C}_N^{1/2} z^N \right]\tag{A.8}$$

and

$$h_{\mathbf{p}}^N(x) := \mathbb{E}_x \left(1 \wedge e^{Q^N} \right).$$

While $h_{\mathbf{p}}$ (see (5.28)) is the limiting average acceptance probability, $h_{\mathbf{p}}^N(x)$ is the *local* average acceptance probability. The above notation will be used in the proof of the next lemma.

Lemma A.2. *If Assumption 5.1, Assumption 5.2 and Condition 5.3 hold, then $w_{\mathbf{p}}^N(t)$ converges weakly in $C([0, T]; \mathcal{H})$ to $D_{\mathbf{p}}W^{\mathcal{C}}(t)$, where $W^{\mathcal{C}}(t)$ is a \mathcal{H} -valued \mathcal{C} -Brownian motion and the constant $D_{\mathbf{p}}$ has been defined in the statement of Theorem 4.2.*

Proof of Lemma A.1. We start by proving (A.7), which is simpler. The drift coefficient $d_{\mathbf{p}}$ is globally Lipschitz; therefore, using (3.6), (3.3) and (3.4), if $t_k \leq t < t_{k+1}$, we have

$$\begin{aligned} \mathbb{E}_{\pi^N} \|d_{\mathbf{p}}(\bar{x}^{(N)}(t)) - d_{\mathbf{p}}(x^{(N)}(t))\| &\lesssim \mathbb{E}_{\pi^N} \|\bar{x}^{(N)}(t) - x^{(N)}(t)\| \\ &\lesssim |(N^{\zeta\gamma}t - k)| \mathbb{E}_{\pi^N} \|x_{k+1}^N - x_k^N\| \lesssim \mathbb{E}_{\pi^N} \|y_{k+1}^N - x_k^N\| \\ &\lesssim \left(\frac{1}{N^{2\gamma}} + \frac{1}{N^{\alpha\gamma}} \right) \mathbb{E}_{\pi^N} \|x_k^N\| + \frac{1}{N^{\gamma}} \mathbb{E} \|(\mathcal{C}^N)^{1/2} z_{k+1}^N\| \rightarrow 0. \end{aligned}$$

Let us now come to the proof of (A.6). From (4.3)-(4.4), we have

$$d_{\mathbf{p}}^N(x) - d_{\mathbf{p}}(x) = A_1^N + A_2^N + A_3^N - d_{\mathbf{p}}(x),$$

where

$$A_1^N := N^{\zeta\gamma} \mathbb{E}_x \left[(1 \wedge e^{Q^N}) \left(-\frac{\ell^2}{2N^{2\gamma}} x^N \right) \right] \quad (\text{A.9})$$

$$\begin{aligned} A_2^N &:= N^{(\zeta-\alpha)\gamma} \ell^\alpha \mathbb{E}_x \left[(1 \wedge e^{Q^N}) \tilde{S}^N x^N \right] \\ A_3^N &:= N^{(\zeta-1)\gamma} \ell \mathbb{E}_x \left[(1 \wedge e^{Q^N}) \mathcal{C}_N^{1/2} z^N \right] \stackrel{(\text{B.17})}{=} N^{(\zeta-1)\gamma} \ell \epsilon_{\mathbf{p}}^N(x). \end{aligned} \quad (\text{A.10})$$

We split the function $d_{\mathbf{p}}(x)$ in three (corresponding) parts:

$$d_{\mathbf{p}}(x) = A_1(x) + A_2(x) + A_3(x),$$

with

$$\begin{aligned} A_1 &:= \begin{cases} -\frac{\ell^2}{2} h_{\mathbf{p}} x & \text{if } \gamma = 1/6 \text{ and } \alpha \geq 2 \\ 0 & \text{otherwise} \end{cases} \\ A_2 &:= \begin{cases} h_{\mathbf{p}} \ell^\alpha \tilde{S} x & \text{if } \gamma \geq 1/6 \text{ and } \alpha \leq 2 \\ 0 & \text{otherwise} \end{cases} \\ A_3 &:= \begin{cases} -2\nu_{\mathbf{p}} \ell^\alpha \tilde{S} x & \text{if } \gamma \geq 1/6 \text{ and } 1 \leq \alpha \leq 2 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

We therefore need to consecutively estimate the above three terms.

- $\mathbf{A}_1^N - \mathbf{A}_1$: if $\alpha \geq 2$ (and $\gamma = 1/6$) we fix $\zeta = 2$ and we have

$$\begin{aligned} \mathbb{E}_{\pi^N} \|A_1^N - A_1\| &\lesssim \mathbb{E}_{\pi^N} \left[\left| \mathbb{E}_x \left(1 \wedge e^{Q^N} \right) - h_{\mathbf{p}} \right| \|x\| \right] + \mathbb{E}_{\pi^N} \|x^N - x\| \\ &\stackrel{(\text{A.1})}{\lesssim} \left(\mathbb{E}_{\pi^N} |h_{\mathbf{p}}^N(x) - h_{\mathbf{p}}|^2 \right)^{1/2} + \mathbb{E}_{\pi^N} \|x^N - x\| \rightarrow 0, \end{aligned} \quad (\text{A.11})$$

as the first addend tends to zero by Lemma B.2 of B and the second addend tends to zero by definition (see also [22, equation (4.3)]). If $\alpha < 2$ then $\zeta = \alpha$ and we have

$$\mathbb{E}_{\pi^N} \|A_1^N\| \lesssim N^{(\alpha-2)\gamma} \mathbb{E}_{\pi^N} \|\mathbb{E}_x(1 \wedge e^{Q^N}) x^N\| \lesssim N^{(\alpha-2)\gamma} \rightarrow 0. \quad (\text{A.12})$$

- $\mathbf{A}_2^N - \mathbf{A}_2$: if $\alpha \leq 2$ then, recalling (4.5), a calculation analogous to the one in (A.11) gives the statement. If $\alpha > 2$ then we can act as in (A.12).

- $\mathbf{A}_3^N - \mathbf{A}_3$: by Lemma B.2 of B (and (4.5)), we have

$$\mathbb{E}_{\pi^N} \|A_3^N - A_3\| \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

This concludes the proof. \square

Proof of Lemma A.2. The calculations here are standard so we only prove it for the case $\gamma = 1/6$ and $\alpha > 2$. Let us recall the martingale difference given by (5.35). In the case $\gamma = 1/6$ and $\alpha > 2$, we have $\zeta = 2$ and $d_{\mathbf{p}}(x) = -\frac{\ell^2}{2} h_{\mathbf{p}} x$. Hence, the expression (5.35) becomes

$$\begin{aligned} M_k^N &= N^{-1/6} \left(\tilde{\beta}^N \frac{1}{h_{\mathbf{p}}} d_{\mathbf{p}}(x_k^N) \right) + N^{(1-\alpha)/6} \left(\ell^\alpha \tilde{\beta}^N \tilde{S} x_k^N \right) + \left(\ell \tilde{\beta}^N \mathcal{C}^{1/2} z_{k+1}^N \right) - N^{-1/6} d_{\mathbf{p}}^N(x_k^N) \\ &= \left(\ell \tilde{\beta}^N \mathcal{C}^{1/2} z_{k+1}^N \right) + N^{-(\alpha-1)/6} \left(\ell^\alpha \tilde{\beta}^N \tilde{S} x_k^N \right) - N^{-1/6} \frac{1}{h_{\mathbf{p}}} \left[h_{\mathbf{p}} d_{\mathbf{p}}^N(x_k^N) - \tilde{\beta}^N d_{\mathbf{p}}(x_k^N) \right] \end{aligned}$$

By Lemma A.1, we have that $\mathbb{E}_{\pi^N} \|d_{\mathbf{p}}^N(x) - d_{\mathbf{p}}(x)\|^2 \rightarrow 0$ as $N \rightarrow \infty$. This implies that

$$N^{-1/3} \mathbb{E}_{\pi^N} \|h_{\mathbf{p}} d_{\mathbf{p}}^N(x) - \tilde{\beta}^N d_{\mathbf{p}}(x)\|^2 \rightarrow 0.$$

At the same time, we also notice that

$$N^{-(\alpha-1)/3} \ell^{2\alpha} \mathbb{E}_{\pi^N} \|\tilde{\beta}^N \tilde{S} x\|^2 \lesssim N^{-(\alpha-1)/3} \ell^{2\alpha} \mathbb{E}_{\pi^N} \|x\|^2 \rightarrow 0 \quad \text{if } \alpha > 1.$$

Hence, if we define $\mathcal{M}_N(x) = \mathbb{E}_x [M_k^N \otimes M_k^N | x_k^N = x]$, we obtain that up to a constant

$$\mathbb{E}_{\pi^N} \left| \text{Tr}(\mathcal{M}_N(x)) - \mathbb{E}_x \left[\|\ell \tilde{\beta}^N \mathcal{C}^{1/2} z\|^2 \right] \right| \leq N^{-1/3}$$

Then, as in Lemma 4.8 of [22] we obtain that

$$\mathbb{E}_{\pi^N} \left| \ell^2 h_{\mathbf{p}} \text{Tr}(\mathcal{C}) - \mathbb{E}_x \left[\|\ell \tilde{\beta}^N \mathcal{C}^{1/2} z\|^2 \right] \right| \rightarrow 0, \quad \text{as } N \rightarrow \infty$$

which then immediately implies that

$$\mathbb{E}_{\pi^N} \left| \ell^2 h_{\mathbf{p}} \text{Tr}(\mathcal{C}) - \text{Tr}(\mathcal{M}_N(x)) \right| \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

The latter result implies that the invariance principle of Proposition 5.1 of [1] holds, which then imply the statement of the lemma. \square

APPENDIX B. AUXILIARY ESTIMATES

We first decompose Q^N as follows: let

$$\begin{aligned} Z^N &:= -\frac{\ell^6}{32} - a - \frac{\ell^3}{4N^{3\gamma}} \langle x, \mathcal{C}^{1/2} z^N \rangle_{\mathcal{C}} \\ &\quad - 2 \frac{\ell^{\alpha-1}}{N^{(\alpha-1)\gamma}} \langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle - \frac{\ell^{(2\alpha-1)}}{N^{(2\alpha-1)\gamma}} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N} \\ &\quad - \frac{\ell^{(3\alpha-1)}}{N^{(3\alpha-1)\gamma}} \langle \tilde{S}^N (\mathcal{C}^N)^{1/2} z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N}. \end{aligned} \tag{B.1}$$

Then

$$Q^N = Z^N + e_{\star}^N, \tag{B.2}$$

where

$$\begin{aligned}
e_\star^N &:= \frac{\ell^6}{32} - \frac{\ell^6}{32 N^{6\gamma}} \|x^N\|_{\mathcal{C}^N}^2 \\
&+ a - 2 \frac{\ell^{2(\alpha-1)}}{N^{(\alpha-1)2\gamma}} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 - \frac{\ell^{2(2\alpha-1)}}{2N^{2\gamma(2\alpha-1)}} \|(\tilde{S}^N)^2 x^N\|_{\mathcal{C}^N}^2 \\
&+ i^N(x, z) + e^N(x, z),
\end{aligned} \tag{B.3}$$

with

$$i^N(x, z) := i_1^N(x, z) + i_2^N(x, z),$$

having defined

$$i_1^N(x, z) := \frac{\ell^4}{8N^{4\gamma}} (\|x^N\|_{\mathcal{C}^N}^2 - \|z^N\|^2) \tag{B.4}$$

$$i_2^N(x, z) := \frac{\ell^{2\alpha}}{2N^{2\alpha\gamma}} \left(\|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2 - \|\tilde{S}^N (\mathcal{C}^N)^{1/2} z^N\|_{\mathcal{C}^N}^2 \right), \tag{B.5}$$

and

$$\begin{aligned}
e^N &:= \frac{\ell^5}{8N^{5\gamma}} \langle x^N, (\mathcal{C}^N)^{1/2} z^N \rangle_{\mathcal{C}^N} - \frac{\ell^{2(\alpha+1)}}{4N^{2(1+\alpha)\gamma}} \|\tilde{S}^N x^N\|_{\mathcal{C}}^2 \\
&- \frac{\ell^{3+\alpha}}{4N^{(3+\alpha)\gamma}} \langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle + \frac{\ell^{(2\alpha+1)}}{2N^{(2\alpha+1)\gamma}} \langle \tilde{S}(\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}^N}.
\end{aligned} \tag{B.6}$$

Finally, we set

$$\tilde{e}^N(x, z) := e^N + \frac{\ell^{2(\alpha+1)}}{4N^{2(1+\alpha)\gamma}} \|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2. \tag{B.7}$$

That is, \tilde{e}^N contains only the addends of e^N that depend on the noise z .

Furthermore, we split $Q^N(x, z)$ into the terms that contain $z^{j,N}$ and the terms that don't, Q_j^N and $Q_{j,\perp}^N$, respectively; that is

$$Q^N = Q_j^N + Q_{j,\perp}^N,$$

where

$$Q_j^N := \tilde{e}^N + (i_1^N)_j + (i_2^N)_j + Z_j^N, \tag{B.8}$$

having denoted by $(i_1^N)_j$, $(i_2^N)_j$ and Z_j^N , the part of i_1^N , i_2^N and Z^N , respectively, that depend on $z^{j,N}$.

Lemma B.1. *Let Assumption 5.1, Assumption 5.2 and Condition 5.3 hold; then,*

$$i) \quad \mathbb{E}_{\pi^N} |\tilde{e}^N|^2 \lesssim \frac{1}{N^{2/3}} + \frac{1}{N^{4\gamma}}, \quad \text{for all } \alpha \geq 1, \gamma \geq 1/6 \quad (\text{B.9})$$

$$ii) \quad N^{1/3} \mathbb{E}_{\pi^N} \sum_{j=1}^N \lambda_j^2 |(i_1^N)_j|^2 \longrightarrow 0, \quad \text{as } N \rightarrow \infty$$

$$iii) \quad \mathbb{E}_{\pi^N} |i_2^N|^2 \lesssim \frac{1}{N^{4\gamma}}, \quad \text{for all } \alpha \geq 1, \gamma \geq 1/6$$

$$iv) \quad N^{1/3} \mathbb{E}_{\pi^N} \sum_{j=1}^N \lambda_j^2 |Z_j^N|^2 \longrightarrow 0, \quad \text{as } N \rightarrow \infty, \text{ for all } \alpha > 2, \gamma = 1/6 \quad (\text{B.10})$$

$$v) \quad \mathbb{E}_{\pi^N} |e^N|^2 \lesssim \frac{1}{N^{2/3}} + \frac{1}{N^{4\gamma}}, \quad \text{for all } \alpha \geq 1, \gamma \geq 1/6 \quad (\text{B.11})$$

$$vi) \quad \mathbb{E}_{\pi^N} |(\text{Var}_x Z^N)^{1/2} - (\text{Var} Q)^{1/2}|^2 \longrightarrow 0, \quad \text{as } N \rightarrow \infty, \text{ for all } \alpha \geq 1, \gamma \geq 1/6. \quad (\text{B.12})$$

Proof of Lemma B.1. Recall that, under π^N , $x^{i,N} \sim \lambda_i \rho^i$, where $\{\rho^i\}_{i \in \mathbb{N}}$ are i.i.d standard Gaussians. We now consecutively prove all the statements of the lemma.

Proof of i). Notice that

$$\mathbb{E}_{\pi^N} |\langle x^N, (\mathcal{C}^N)^{1/2} z^N \rangle_{\mathcal{C}^N}|^2 = \mathbb{E} \left| \sum_{i=1}^N \rho^i z^{i,N} \right|^2 = N.$$

Therefore, since $\gamma \geq 1/6$,

$$\mathbb{E}_{\pi^N} \left| \frac{\ell^5}{8N^{5\gamma}} \langle x^N, (\mathcal{C}^N)^{1/2} z^N \rangle_{\mathcal{C}^N} \right|^2 \lesssim N^{-2/3}. \quad (\text{B.13})$$

Furthermore, using (5.15) (which follows from point i) of Assumption 5.2) and (5.11), we have

$$\frac{1}{N^{(3+\alpha)2\gamma}} \mathbb{E}_{\pi^N} |\langle (\mathcal{C}^N)^{1/2} z^N, S^N x^N \rangle|^2 \lesssim N^{-8\gamma}; \quad (\text{B.14})$$

similarly, using (5.16) (which follows from point ii) of Assumption 5.2) and (5.12),

$$\frac{1}{N^{(2\alpha+1)2\gamma}} \mathbb{E}_{\pi^N} |\langle \tilde{S}(\mathcal{C}^N)^{1/2} z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}}|^2 \lesssim N^{-4\gamma}. \quad (\text{B.15})$$

Now the first statement of the lemma is a consequence of (B.6)-(B.7) and the above (B.13), (B.14) and (B.15).

Proof of ii). This is proved in [22], see calculations after [22, equation (4.18)], so we omit it.

Proof of iii). This estimate follows again from Assumption 5.2, once we observe that if $x \sim \pi^N$ then $\|\tilde{S}^N x^N\|_{\mathcal{C}^N}^2$ and $\|\tilde{S}^N (\mathcal{C}^N)^{1/2} z^N\|_{\mathcal{C}^N}^2$ are two independent random variables with

the same distribution. With this observation in place, we have

$$\begin{aligned} \mathbb{E}_{\pi^N} \left| \frac{\|\tilde{S}x^N\|_{\mathcal{C}}^2 - \|\tilde{S}\mathcal{C}_N^{1/2}z^N\|_{\mathcal{C}}^2}{N^{2\alpha\gamma}} \right|^2 &= \mathbb{E}_{\pi^N} \left| \frac{\|\tilde{S}x^N\|_{\mathcal{C}}^2}{N^{2\alpha\gamma}} - \frac{c_1}{N^{2\gamma}} + \frac{c_1}{N^{2\gamma}} - \frac{\|\tilde{S}\mathcal{C}_N^{1/2}z^N\|_{\mathcal{C}}^2}{N^{2\alpha\gamma}} \right|^2 \\ &\lesssim \mathbb{E}_{\pi^N} \left| \frac{\|\tilde{S}x^N\|_{\mathcal{C}}^2}{N^{2\alpha\gamma}} - \frac{c_1}{N^{2\gamma}} \right|^2 = N^{-4\gamma} \mathbb{E}_{\pi^N} \left| \frac{\|\tilde{S}x^N\|_{\mathcal{C}}^2}{N^{2(\alpha-1)\gamma}} - c_1 \right|^2, \end{aligned}$$

which gives the claim.

Proof of iv). We recall that Z_j^N has been introduced in (B.8). Using the antisymmetry of S^N and the definition of \tilde{S}^N , we have

$$-\langle \tilde{S}^N(\mathcal{C}^N)^{1/2}z^N, \tilde{S}^N x^N \rangle_{\mathcal{C}} = \langle (\mathcal{C}^N)^{1/2}z^N, S^N \tilde{S}^N x^N \rangle$$

and

$$-\langle \tilde{S}^N(\mathcal{C}^N)^{1/2}z^N, (\tilde{S}^N)^2 x^N \rangle_{\mathcal{C}^N} = \langle (\mathcal{C}^N)^{1/2}z^N, S^N (\tilde{S}^N)^2 x^N \rangle.$$

We can therefore write an explicit expression for Z_j^N :

$$\begin{aligned} Z_j^N &= -\frac{\ell^3}{4N^{3\gamma}} \frac{x^{j,N} z^{j,N}}{\lambda_j} - 2 \frac{\ell^{\alpha-1}}{N^{(\alpha-1)\gamma}} \lambda_j z^{j,N} (S^N x^N)^j \\ &\quad + \frac{\ell^{2\alpha-1}}{N^{(2\alpha-1)\gamma}} \lambda_j z^{j,N} (S^N \tilde{S}^N x^N)^j + \frac{\ell^{3\alpha-1}}{N^{(3\alpha-1)\gamma}} \lambda_j z^{j,N} (S^N (\tilde{S}^N)^2 x^N)^j. \end{aligned} \quad (\text{B.16})$$

For the sake of clarity we stress again that in the above $(S^N x^N)$ is an N -dimensional vector and $(S^N x^N)^j$ is the j -th component of such a vector. Therefore, recalling (2.3), (2.1), (A.1) and setting $\gamma = 1/6$, we have

$$\begin{aligned} \sum_{j=1}^N \lambda_j^2 \mathbb{E}_{\pi^N} |Z_j^N|^2 &\lesssim \frac{1}{N} \sum_{j=1}^N \lambda_j^2 \mathbb{E} |\rho^j z^{j,N}|^2 + \frac{\mathbb{E}_{\pi^N}}{N^{(\alpha-1)/3}} \sum_{j=1}^N \lambda_j^4 |(S^N x^N)^j|^2 \\ &\quad + \frac{\mathbb{E}_{\pi^N}}{N^{(2\alpha-1)/3}} \sum_{j=1}^N \lambda_j^4 |(S^N \tilde{S}^N x^N)^j|^2 + \frac{\mathbb{E}_{\pi^N}}{N^{(3\alpha-1)/3}} \sum_{j=1}^N \lambda_j^4 |(S^N (\tilde{S}^N)^2 x^N)^j|^2 \\ &\lesssim \frac{1}{N} + \frac{1}{N^{(\alpha-1)/3}} \mathbb{E}_{\pi^N} \|\tilde{S}^N x^N\|^2 + \frac{1}{N^{(2\alpha-1)/3}} \mathbb{E}_{\pi^N} \|(\tilde{S}^N)^2 x^N\|^2 \\ &\quad + \frac{1}{N^{(3\alpha-1)/3}} \mathbb{E}_{\pi^N} \|(\tilde{S}^N)^3 x^N\|^2 \\ &\lesssim \frac{1}{N} + \frac{1}{N^{(\alpha-1)/3}} \mathbb{E}_{\pi^N} \|x^N\|^2. \end{aligned}$$

Therefore,

$$N^{1/3} \sum_{j=1}^N \lambda_j^2 \mathbb{E}_{\pi^N} |Z_j^N|^2 \lesssim \frac{1}{N^{2/3}} + \frac{1}{N^{(\alpha-2)/3}} \longrightarrow 0, \quad \text{when } \alpha > 2.$$

Proof of v). Follows from Assumption 5.2, from statement i) of this lemma and from (B.7).

Proof of vi). From (B.1),

$$\begin{aligned} \text{Var}_x(Z^N) = & \mathbb{E}_x \left| -\frac{\ell^3}{4N^{3\gamma}} \langle x, \mathcal{C}^{1/2} z^N \rangle_{\mathcal{C}} - 2 \frac{\ell^{\alpha-1}}{N^{(\alpha-1)\gamma}} \langle \mathcal{C}^{1/2} z^N, Sx \rangle \right. \\ & \left. - \frac{\ell^{(2\alpha-1)}}{N^{(2\alpha-1)\gamma}} \langle \tilde{S} \mathcal{C}_N^{1/2} z^N, \tilde{S}x \rangle_{\mathcal{C}} - \frac{\ell^{(3\alpha-1)}}{N^{(3\alpha-1)\gamma}} \langle \tilde{S} \mathcal{C}_N^{1/2} z^N, \tilde{S}^2 x \rangle_{\mathcal{C}} \right|^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}_x(Z^N) = & \frac{\ell^6}{16N^{6\gamma}} \mathbb{E}_x \|x\|_{\mathcal{C}}^2 + 4 \frac{\ell^{2(\alpha-1)}}{N^{2(\alpha-1)\gamma}} \mathbb{E}_x \|\mathcal{C}_N^{1/2} Sx\|^2 \\ & + \frac{\ell^{2(2\alpha-1)}}{N^{2(2\alpha-1)\gamma}} \mathbb{E}_x \|\mathcal{C}_N^{1/2} S \tilde{S}x\|^2 + \frac{\ell^{2(3\alpha-1)}}{N^{2(3\alpha-1)\gamma}} \mathbb{E}_x \|\mathcal{C}_N^{1/2} S \tilde{S}^2 x\|^2 + \mathbb{E}_x r^N \end{aligned}$$

where r^N contains all the cross-products in the expansion of the variance. By direct calculation and using the antisymmetry of S , one finds that most of such cross products vanish and we have

$$\mathbb{E}_x r^N := \frac{\ell^{(2\alpha+2)}}{2N^{3\gamma} N^{(2\alpha-1)\gamma}} \langle Sx, \tilde{S}x \rangle + 4 \frac{\ell^{(4\alpha-2)}}{N^{2(2\alpha-1)\gamma}} \mathbb{E}_x \|\tilde{S}^2 x\|_{\mathcal{C}}^2.$$

Observe that

$$\langle Sx, \tilde{S}x \rangle = \|\tilde{S}x\|_{\mathcal{C}}^2;$$

using this fact, Assumption 5.2 implies that the first addend in the above expression for $\mathbb{E}_x r^N$ vanishes as $N \rightarrow \infty$. The second addend contributes instead to the limiting variance. Now straightforward calculations give the result. \square

We recall the definitions

$$\epsilon_{\mathbf{p}}^N(x) := \mathbb{E}_x \left[\left(1 \wedge e^{Q^N} \right) \mathcal{C}_N^{1/2} z^N \right] \quad (\text{B.17})$$

and

$$h_{\mathbf{p}}^N(x) := \mathbb{E}_x \left(1 \wedge e^{Q^N} \right).$$

Lemma B.2. *Suppose that Assumption 5.1, Assumption 5.2 and Condition 5.3 hold. Then*

(i) *If $\alpha > 2$ and $\gamma = 1/6$,*

$$N^{1/3} \mathbb{E}_{\pi^N} \|\epsilon_{\mathbf{p}}^N(x)\|^2 \xrightarrow{N \rightarrow \infty} 0; \quad (\text{B.18})$$

(ii) *if $1 \leq \alpha \leq 2$ and $\gamma \geq 1/6$ then*

$$\mathbb{E}_{\pi^N} \|N^{\gamma(\alpha-1)} \epsilon_{\mathbf{p}}^N(x) + 2\ell^{\alpha-1} \nu_{\mathbf{p}} \tilde{S}x\|^2 \xrightarrow{N \rightarrow \infty} 0 \quad (\text{B.19})$$

where the constant $\nu_{\mathbf{p}}$ has been defined in (5.31).

(iii) *if $\alpha \geq 1$, $\gamma \geq 1/6$ and S^N is such that $(c_1, c_2, c_3) \neq (0, 0, 0)$, then*

$$\mathbb{E}_{\pi^N} |h_{\mathbf{p}}^N(x) - h_{\mathbf{p}}|^2 \xrightarrow{N \rightarrow \infty} 0; \quad (\text{B.20})$$

(iv) *finally, if $1 < \alpha < 2$, $\gamma > 1/6$ and S^N is such that $c_1 = c_2 = c_3 = 0$, then*

$$\mathbb{E}_{\pi^N} |h_{\mathbf{p}}^N(x) - 1|^2 \xrightarrow{N \rightarrow \infty} 0,$$

i.e. the constant $h_{\mathbf{p}}$ in (B.20) is equal to one. This means that, as $N \rightarrow \infty$, the acceptance probability tends to one.

Proof of Lemma B.2. • Proof of (i). Acting as in [22, page 2349], we obtain

$$|\langle \epsilon_{\mathbf{p}}^N(x), \varphi_j \rangle|^2 \lesssim \lambda_j^2 \mathbb{E}_x |Q_j^N|^2.$$

Taking the sum over j on both sides of the above then gives

$$\|\epsilon_{\mathbf{p}}^N(x)\|^2 \lesssim \sum_{j=1}^N \lambda_j^2 \mathbb{E}_x |Q_j^N|^2.$$

Therefore, if we show

$$N^{1/3} \sum_{j=1}^N \lambda_j^2 \mathbb{E}_x |Q_j^N|^2 \xrightarrow{N \rightarrow \infty} 0,$$

(B.18) follows. From (B.8), it is clear that the above is a consequence of Lemma B.1 (in particular, it follows from (B.9)-(B.10)).

• **Proof of (ii).** Let us split Q^N as follows:

$$Q^N = R^N + e^N + i_2^N,$$

where e^N and i_2^N are defined in (B.6) and (B.5), respectively, while

$$R^N := I^N + i_1 + B^N + H^N, \quad (\text{B.21})$$

having set

$$I^N := -\frac{\ell^6}{32 N^{6\gamma}} \|x^N\|_{\mathcal{C}}^2 - 2 \frac{\ell^{2(\alpha-1)}}{N^{2\gamma(\alpha-1)}} \|\tilde{S}x^N\|_{\mathcal{C}}^2 - \frac{\ell^{2(2\alpha-1)}}{2N^{2\gamma(2\alpha-1)}} \|\tilde{S}^2x^N\|_{\mathcal{C}}^2 \quad (\text{B.22})$$

$$B^N := -2 \frac{\ell^{\alpha-1}}{N^{\gamma(\alpha-1)}} \langle \mathcal{C}^{1/2} z^N, Sx^N \rangle \quad (\text{B.23})$$

$$\begin{aligned} H^N &:= -\frac{\ell^3}{4N^{3\gamma}} \langle x, \mathcal{C}^{1/2} z^N \rangle_{\mathcal{C}} - \frac{\ell^{(2\alpha-1)}}{N^{\gamma(2\alpha-1)}} \langle \tilde{S} \mathcal{C}_N^{1/2} z^N, \tilde{S}x^N \rangle_{\mathcal{C}} \\ &\quad - \frac{\ell^{(3\alpha-1)}}{N^{\gamma(3\alpha-1)}} \langle \tilde{S} \mathcal{C}_N^{1/2} z^N, \tilde{S}^2x^N \rangle_{\mathcal{C}}, \end{aligned} \quad (\text{B.24})$$

and i_1^N is defined in (B.4). The j -th component of $N^{\gamma(\alpha-1)} \epsilon_{\mathbf{p}}^N$ can be therefore expressed as follows:

$$N^{\gamma(\alpha-1)} \epsilon_{\mathbf{p}}^{j,N} = N^{\gamma(\alpha-1)} \mathbb{E}_x \left[(1 \wedge e^{Q^N}) \lambda_j z^{j,N} \right] = N^{\gamma(\alpha-1)} \mathbb{E}_x \left[(1 \wedge e^{R^N}) \lambda_j z^{j,N} \right] + T_0^j \quad (\text{B.25})$$

where $T_0^j := \langle T_0, \varphi_j \rangle$ and

$$T_0 := N^{\gamma(\alpha-1)} \mathbb{E}_x \left[\left((1 \wedge e^{Q^N}) - (1 \wedge e^{R^N}) \right) \mathcal{C}_N^{1/2} z^N \right]. \quad (\text{B.26})$$

We now decompose R^N into a component which depends on $z^{j,N}$, R_j^N , and a component that does not depend on $z^{j,N}$, $R_{j,\perp}^N$:

$$R^N = R_j^N + R_{j,\perp}^N,$$

with

$$R_j^N := (i_1)_j + (B^N)_j + (H^N)_j,$$

having denoted by $(i_1^N)_j$, $(B^N)_j$ and $(H^N)_j$ the part of i_1 , B^N and H^N , respectively, that depend on $z^{j,N}$. That is,

$$(i_1)_j = -\frac{\ell^4}{8N^{4\gamma}} |z^{j,N}|^2;$$

as for $(H^N)_j$, it suffices to notice that

$$(B^N)_j + (H^N)_j = Z_j^N,$$

and the expression for Z_j^N is detailed in (B.16) (just set $\alpha = 2$ in (B.16)). With this notation, from (B.25), we further write

$$N^{\gamma(\alpha-1)} \mathbb{E}_x \left[(1 \wedge e^{Q^N}) \lambda_j z^j \right] = N^{\gamma(\alpha-1)} \mathbb{E}_x \left[(1 \wedge e^{[R^N - (i_1^N)_j - H_j^N]}) \lambda_j z^j \right] + T_0^j + T_1^j \quad (\text{B.27})$$

where, like before, $T_1^j := \langle T_1, \varphi_j \rangle$ and

$$T_1 := N^{\gamma(\alpha-1)} \ell \mathbb{E}_x \left[\left((1 \wedge e^{R^N}) - \left(1 \wedge e^{R^N - (i_1^N)_j - H_j^N} \right) \right) \mathcal{C}_N^{1/2} z^N \right].$$

We recall that the notation \mathbb{E}_x denotes expected value given x , where the expectation is taken over all the sources of noise contained in the integrand. In order to further evaluate the RHS of (B.27) we calculate the expected value of the integrand with respect to the law of z^j (we denote such expected value by \mathbb{E}^{z^j} and use \mathbb{E}^{z^j} to denote expectation with respect to $z^N \setminus z^{j,N}$); to this end, we use the following lemma.

Lemma B.3. *If G is a normally distributed random variable with $G \sim \mathcal{N}(0, 1)$ then*

$$\mathbb{E} [G (1 \wedge e^{\delta G + \mu})] = \delta e^{\mu + \delta^2/2} \Phi \left(-\frac{\mu}{|\delta|} - |\delta| \right),$$

where Φ is the CDF of a standard Gaussian.

We apply the above lemma with $\mu = R_{j,\perp}^N$ and $\delta = \delta_j^B$, where

$$\delta_j^B := -2 \frac{\ell^{\alpha-1}}{N^{\gamma(\alpha-1)}} \lambda_j (Sx^N)^j.$$

We therefore obtain

$$\begin{aligned} N^{\gamma(\alpha-1)} \lambda_j \mathbb{E}_x \left[(1 \wedge e^{[R^N - (i_1^N)_j - H_j^N]}) z^j \right] &= N^{\gamma(\alpha-1)} \lambda_j \mathbb{E}_x^{z^j} \delta_j^B e^{R_{j,\perp}^N + (\delta_j^B)^2/2} \Phi \left(-\frac{R_{j,\perp}^N}{|\delta_j^B|} - |\delta_j^B| \right) \\ &= -2\ell^{\alpha-1} \lambda_j^2 (S^N x^N)^j \mathbb{E}_x^z e^{R_{j,\perp}^N + (\delta_j^B)^2/2} \Phi \left(-\frac{R_{j,\perp}^N}{|\delta_j^B|} - |\delta_j^B| \right) \\ &= -2\ell^{\alpha-1} \lambda_j^2 \mathbb{E}_x (Sx)^j e^{R_{j,\perp}^N + (\delta_j^B)^2/2} \mathbf{1}_{\{R_{j,\perp}^N < 0\}} + T_2^j + T_3^j \\ &= -2\ell^{\alpha-1} \lambda_j^2 \mathbb{E}_x (Sx)^j e^{\tilde{Q}^N} \mathbf{1}_{\{\tilde{Q} < 0\}} + T_2^j + T_3^j + T_4^j \\ &= -2\ell^{\alpha-1} \lambda_j^2 \mathbb{E}_x (Sx)^j e^{Q^N} \mathbf{1}_{\{Q < 0\}} + T_2^j + T_3^j + T_4^j + T_5^j, \end{aligned}$$

$$T_2^j := -2\ell^{\alpha-1} \lambda_j^2 \mathbb{E}_x (Sx)^j e^{R_{j,\perp}^N + (\delta_j^B)^2/2} \left[\Phi \left(-\frac{R_{j,\perp}^N}{|\delta_j^B|} - |\delta_j^B| \right) - \Phi \left(-\frac{R_{j,\perp}^N}{|\delta_j^B|} \right) \right]$$

and

$$\begin{aligned}
T_3^j &:= -2\ell^{\alpha-1}\lambda_j^2\mathbb{E}_x(Sx)^j e^{R_{j,\perp}^N + (\delta_j^B)^2/2} \left[\Phi\left(-\frac{R_{j,\perp}^N}{|\delta_j^B|}\right) - \mathbf{1}_{\{R_{j,\perp}^N < 0\}} \right] \\
T_4^j &:= -2\ell^{\alpha-1}\lambda_j^2\mathbb{E}_x(Sx)^j \left[e^{R_{j,\perp}^N + (\delta_j^B)^2/2} \mathbf{1}_{\{R_{j,\perp}^N < 0\}} - e^{Q^N} \mathbf{1}_{\{Q^N < 0\}} \right] \\
T_5^j &:= -2\ell^{\alpha-1}\lambda_j^2\mathbb{E}_x(Sx)^j \left[e^{Q^N} \mathbf{1}_{\{Q^N < 0\}} - e^Q \mathbf{1}_{\{Q < 0\}} \right].
\end{aligned} \tag{B.28}$$

To prove the statement it suffices to show that

$$\mathbb{E}_{\pi^N} \sum_{n=0}^5 \|T_n\|^2 \rightarrow 0 \quad \text{as } N \rightarrow \infty$$

These calculations are a bit lengthy, so we gather the proof of the above in Lemma B.4 below. Assuming for the moment that the above is true, the proof is concluded after recognising that

$$-2\ell^2\lambda_j^2(Sx)^j \mathbb{E} e^Q \mathbf{1}_{\{Q < 0\}} = -2\ell^2(\tilde{S}^N x^N)^j \nu_{\mathbf{p}}.$$

• **Proof of (iii).** By acting as in the proof of [22, Lemma 4.5 and Corollary 4.6] we see that (B.20) is a consequence of (B.2), (B.12) and of the following limit:

$$\mathbb{E}_{\pi^N} |e_\star^N|^2 \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

The above follows from the definition (B.3), Lemma B.1, Assumption 5.2 and [22, equation (4.7)].

• **Proof of (iv).** One could show this with the same procedure as in (iii). However, in this case things are easier, indeed we can write

$$\mathbb{E}_{\pi^N} \left| \mathbb{E}_x(1 \wedge e^{Q^N}) - (1 \wedge e^0) \right|^2 \leq \mathbb{E}_{\pi^N} |Q^N|^2 \rightarrow 0.$$

The above limit follows simply by the assumption that $\gamma > 1/6$ and $c_1 = c_2 = c_3 = 0$. \square

Lemma B.4. *If Assumption 5.1, Assumption 5.2, Assumption 5.4 and Condition 5.3 hold, then*

$$\mathbb{E}_{\pi^N} \|T_n\|^2 = \mathbb{E}_{\pi^N} \sum_{i=1}^N |T_n^j|^2 \xrightarrow{N \rightarrow \infty} 0, \quad \text{for all } n \in \{0, 1, \dots, 5\},$$

where $T_n = \sum_{i=1}^N \langle T_n, \varphi_i \rangle \varphi_i$ and the terms T_0^j, \dots, T_5^j have been introduced in (B.26)- (B.28).

Proof of Lemma B.4. We consecutively prove the above limit for the terms T_0^j, \dots, T_5^j .

• Using the Lipschitzianity of the function $u \rightarrow 1 \wedge e^u$, the result for T_0 follows from the definition of R^N , equation (B.21), and Lemma B.1, statements iii) and v). The result for T_1 can be obtained similarly.

• Term T_2 : we use the Lipschitzianity of the function Φ and observe that the following holds

$$\mathbb{E}_{\pi^N} e^{c(R_{j,\perp}^N + \delta^2/2)} \lesssim \mathbb{E}_{\pi^N} e^{cR^N} \lesssim 1 \quad \text{for all } c > 0.$$

The above can be obtained with a reasoning similar to the one detailed in [18, page 916 and (5.20)], using (i) of Assumption 5.4. Using the above observations and applying the Hölder

inequality with the exponent r appearing in (ii) of Assumption 5.4, one then gets,

$$\mathbb{E}_{\pi^N} \|T_2\|^2 \lesssim \left(\mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \frac{\lambda_j^6 |(Sx)^j|^4}{N^{2\gamma(\alpha-1)}} \right)^r \right)^{1/r}.$$

Therefore the term T_2 goes to zero by Assumption 5.4.

• The term T_3 can be treated with calculations completely analogous to those in [18, Lemma 5.8]. As a result of such calculations, using the fact that the noise $z^{j,N}$ is always independent on the current position x , and recalling equation (B.16), we obtain that for any $r, q > 1$ (to be later appropriately chosen), the following bound holds:

$$\mathbb{E}_{\pi^N} \|T_3\|^2 \lesssim \left\{ \mathbb{E}_{\pi^N} \left[\sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x \frac{(\lambda_j |(Sx)^j| + 1)}{|R^N| N^{\gamma(\alpha-1)} + 1} \cdot \left(1 + \frac{|z^{j,N}|^2}{N^{4\gamma}} + Z_j^N \right) \right]^{2/q} \right]^r \right\}^{1/r} \quad (\text{B.29})$$

Now set

$$D_N := \left(\mathbb{E}_x \frac{1}{(|R^N| N^{\gamma(\alpha-1)} + 1)^2} \right)^{r/q},$$

so that

$$\mathbb{E}_{\pi^N} \|T_3\|^2 \lesssim \left\{ \mathbb{E}_{\pi^N} D_N \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x (\lambda_j |(Sx)^j| + 1)^2 \right]^{1/q} \right\}^r \right\}^{1/r} \quad (I)$$

$$+ \left\{ \mathbb{E}_{\pi^N} D_N \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x (\lambda_j |(Sx)^j| + 1)^2 |Z_j^N|^2 \right]^{1/q} \right\}^r \right\}^{1/r}. \quad (II)$$

Notice that by the bounded convergence theorem, we have

$$\mathbb{E}_{\pi^N} D_N \lesssim \left(\frac{1}{N^{\gamma(\alpha-1)}} \right)^{r/q}. \quad (\text{B.30})$$

With this observation it is easy to show that the term (I) tends to zero. It is less easy to show that (II) tends to zero, so for this term we detail calculations a bit more.

$$\begin{aligned}
(II) &\lesssim (\mathbb{E}_{\pi^N} D_N^2)^{1/(2r)} \left(\mathbb{E}_{\pi^N} \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x \left[(\lambda_j |(Sx)^j| + 1) \frac{|x^{j,N} z^{j,N}|}{N^{3\gamma} \lambda_j} \right]^2 \right]^{1/q} \right\}^{2r} \right)^{1/(2r)} \\
&+ (\mathbb{E}_{\pi^N} D_N^2)^{1/(2r)} \left(\mathbb{E}_{\pi^N} \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x \left[(\lambda_j |(Sx)^j| + 1) \frac{\lambda_j |(Sx)^j| |z^{j,N}|}{N^{\gamma(\alpha-1)}} \right]^2 \right]^{1/q} \right\}^{2r} \right)^{1/(2r)} \\
&+ (\mathbb{E}_{\pi^N} D_N^2)^{1/(2r)} \left(\mathbb{E}_{\pi^N} \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x \left[(\lambda_j |(Sx)^j| + 1) \frac{\lambda_j |(S\tilde{S}x)^j| |z^{j,N}|}{N^{\gamma(2\alpha-1)}} \right]^2 \right]^{1/q} \right\}^{2r} \right)^{1/(2r)} \\
&+ (\mathbb{E}_{\pi^N} D_N^2)^{1/(2r)} \left(\mathbb{E}_{\pi^N} \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x \left[(\lambda_j |(Sx)^j| + 1) \frac{\lambda_j |(S\tilde{S}^2x)^j| |z^{j,N}|}{N^{\gamma(3\alpha-1)}} \right]^2 \right]^{1/q} \right\}^{2r} \right)^{1/(2r)}.
\end{aligned}$$

We denote by $(II)_1$ to $(II)_4$ the terms in line 1 to 4 of the above array of equations and the second factor in line i we denote by $(II)_{ib}$, so e.g.

$$(II)_1 = (\mathbb{E}_{\pi^N} D_N^2)^{1/(2r)} ((II)_{1b})^{1/(2r)},$$

where

$$(II)_{1b} := \mathbb{E}_{\pi^N} \left\{ \sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \left[\mathbb{E}_x \left[(\lambda_j |(Sx)^j| + 1) \frac{|x^{j,N} z^{j,N}|}{N^{3\gamma} \lambda_j} \right]^2 \right]^{1/q} \right\}^{2r}.$$

To streamline the presentation we have written the calculations leading to the above four addends in a way that it looks like the choice of q should be the same for the four terms above. However, acting appropriately in the computations that give (B.29), one can see that the q does not need to be the same for each one of the above addends. We show how to study $(II)_1$ and $(II)_3$, the other terms can be done with similar tricks. Starting from $(II)_1$, because of (B.30), we just need to prove that $(II)_{1b}$ is bounded. We will do slightly better in what follows. Recall that by assumption

$$\mathbb{E}_{\pi^N} \sum_{j=1}^N \frac{\lambda_j^{2p} |(Sx)^j|^{2p}}{N^{2p\gamma(\alpha-1)}} \leq \mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \frac{\lambda_j^2 |(Sx)^j|^2}{N^{2\gamma(\alpha-1)}} \right)^p < \infty. \quad (\text{B.31})$$

Choosing $q = 2$ in the definition of $(II)_{1b}$ and recalling $x^{j,N} \sim \lambda_j \rho_j$, we get

$$(II)_{1b} = \mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \frac{\lambda_j^2 \lambda_j^3 |(Sx)^j|^3}{N^{3\gamma}} \right)^{2r} \lesssim \mathbb{E}_{\pi^N} \sum_{j=1}^N \frac{\lambda_j^2 \lambda_j^{6r} |(Sx)^j|^{6r}}{N^{6r\gamma}} \longrightarrow 0,$$

where in the last inequality we have used the weighted Jentsen's inequality (relying on the fact that $\{\lambda_j^2\}_j$ is summable) and the convergence of the RHS to zero follows from (B.31). The term $(II)_{2b}$ can be dealt with analogously, choosing $q = 4$ (this time when applying the weighted Jentsen's inequality one should rely on the fact that the sequence $\{\lambda_j^4 |(Sx)^j|^2\}_j$ is summable for every $x \in \mathcal{H}$). Finally, to deal with $(II)_{3b}$, we use the fact that the sequence $\{(\tilde{S}^2 x)^j\}_j$ is, by assumption, bounded for every $x \in \mathcal{H}$. Therefore, choosing $q = 2$ we have:

$$\begin{aligned} (II)_{3b} &= \mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \lambda_j^4 |(Sx)^j|^2 \lambda_j |(Sx)^j| \frac{\lambda_j |(S\tilde{S}x)^j|}{N^{\gamma(2\alpha-1)}} \right)^{2r} \\ &= \mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \frac{\lambda_j^4 |(Sx)^j|^3}{N^{\gamma(2\alpha-1)}} |(\tilde{S}^2 x)^j| \right)^{2r} \\ &\leq \mathbb{E}_{\pi^N} \left(\sum_{j=1}^N \frac{\lambda_j^2 |(Sx)^j|^2}{N^{\gamma(2\alpha-1)}} \lambda_j^2 |(Sx)^j| \right)^{2r}. \end{aligned}$$

Because $2\alpha\gamma - \gamma > 2\gamma(\alpha - 1)$, the RHS of the above tends to zero by using (B.31). The term $(II)_{4b}$ can be dealt with in a completely analogous manner.

• The terms T_4 and T_5 can be studied similarly to what has been done in [14], see calculations from equation (8.31), in particular the terms $e_{3,k}^{i,N}, e_{5,k}^{i,N}$. \square

REFERENCES

- [1] Erich Berger. Asymptotic behaviour of a class of stochastic approximation procedures. *Probability theory and related fields*, 71(4):517–552, 1986.
- [2] Joris Bierkens. Non-reversible metropolis-hastings. *Statistics and Computing*, pages 1–16, 2015.
- [3] A. Bouchard-Côte, A. Doucet and S.J. Vollmer. The bouncy particle sampler: A non-reversible rejection-free Markov Chain Monte Carlo method. *submitted*, 2015.
- [4] O. F. Christensen, G. O. Roberts, and J. S. Rosenthal. *Scaling limits for the transient phase of local Metropolis-Hastings algorithms*. J. R. Stat. Soc. Ser. B Stat. Methodol., 67(2):253 – 268, 2005.
- [5] Persi Diaconis, Susan Holmes, and Radford M Neal. Analysis of a nonreversible Markov chain sampler. *Annals of Applied Probability*, pages 726–752, 2000.
- [6] AB Duncan, T Lelievre, and GA Pavliotis. Variance reduction using nonreversible Langevin samplers. *Journal of Statistical Physics*, 163(3):457–491, 2016.
- [7] AB Duncan, GA Pavliotis, and K Zygalakis. Nonreversible Langevin samplers: Splitting schemes, analysis and implementation. *submitted*, 2017.
- [8] Aryeh Dvoretzky et al. Asymptotic normality for sums of dependent random variables. In *Proc. 6th Berkeley Symp. Math. Statist. Probab*, volume 2, pages 513–535, 1972.
- [9] Alan M Horowitz. A generalized guided Monte Carlo algorithm. *Physics Letters B*, 268(2):247–252, 1991.
- [10] Chii-Ruey Hwang, Shu-Yin Hwang-Ma, and Shuenn-Jyi Sheu. Accelerating diffusions. *The Annals of Applied Probability*, 15(2):1433–1444, 2005.
- [11] B. Jourdain, T. Lelievre, and B. Miasojedow. *Optimal scaling for the transient phase of Metropolis Hastings algorithms: the longtime behavior*. Bernoulli, 20(4):1930 – 1978, 2014.
- [12] B. Jourdain, T. Lelievre, and B. Miasojedow. *Optimal scaling for the transient phase of the random walk Metropolis algorithm: the mean field limit*. Ann. Appl. Probab., 25(4):2263 – 2300, 2015.
- [13] E. P. Bernard, W. Krauth, D. B. Wilson. *Event-chain algorithms for hard-sphere systems*. Physical Review E, (2009)

- [14] J. Kuntz, Michela Ottobre, and Andrew M Stuart. Diffusion limit for the random walk metropolis algorithm out of stationarity. *arXiv preprint arXiv:1405.4896*, 2014.
- [15] J. Kuntz, Michela Ottobre, and Andrew M Stuart. *Non-stationary phase of the MALA algorithm*. *arXiv preprint arXiv:1608.08379*, 2016.
- [16] Jianfeng Lu and Konstantinos Spiliopoulos. Multiscale integrators for stochastic differential equations and irreversible langevin samplers. *arXiv preprint arXiv:1606.09539*, 2016.
- [17] Y.-A Ma, T. Chen and E. Fox. *A complete recipe for stochastic-gradient MCMC*. Advances in Neural Information Processing systems 28, pages 2899–2907, 2015.
- [18] Jonathan C Mattingly, Natesh S Pillai, and Andrew M Stuart. Diffusion limits of the random walk metropolis algorithm in high dimensions. *The Annals of Applied Probability*, 22(3):881–930, 2012.
- [19] P. Monmarche. *Piecewise deterministic simulated annealing*, ALEA, Lat. Am. J. Probab. Math. Stat. 13 (1), 357–398 (2016)
- [20] M. Ottobre. *Markov Chain Monte Carlo and Irreversibility*. Reports on Math. Phys. (2016)
- [21] Michela Ottobre, Natesh S Pillai, Frank J Pinski, and Andrew M Stuart. A function space hmc algorithm with second order Langevin diffusion limit. *Bernoulli*, 22(1):60–106, 2016.
- [22] Natesh S Pillai, Andrew M Stuart, and Alexandre H Thiéry. Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *The Annals of Applied Probability*, 22(6):2320–2356, 2012.
- [23] R. Poncet. *Generalized and hybrid Metropolis-Hastings overdamped Langevin algorithms*. arXiv 1701.05833
- [24] Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081–2103, 2015.
- [25] Luc Rey-Bellet and Konstantinos Spiliopoulos. Variance reduction for irreversible Langevin samplers and diffusion on graphs. *Electronic Communications in Probability*, 20, 2015.
- [26] Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.
- [27] Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367, 2001.
- [28] Gareth O Roberts and Richard L Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.
- [29] P.J. Rossky, J.D. Doll and H.L. Friedman, *Brownian Dynamics as smart Monte Carlo Simulation*, J.Chem. Phys., 69 (10), 4628–4633, 1978.

DEPARTMENT OF MATHEMATICS, HERIOT WATT UNIVERSITY, EDINBURGH, EH14 4AS, SCOTLAND
E-mail address: m.ottobre@hw.ac.uk

DEPARTMENT OF STATISTICS, HARVARD UNIVERSITY, CAMBRIDGE, MA, 02138-2901, USA
E-mail address: pillai@fas.harvard.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, BOSTON UNIVERSITY, BOSTON, MA, 02215, USA
E-mail address: kspiliop@math.bu.edu